Engineering Adeno-Associated Virus for Receptor-Mediated Gene Delivery

A DISSERTATION SUBMITTED TO THE GRADUATE FACULTY OF THE UNIVERSITY OF MINNESOTA BY

Alina Catherine Zdechlik

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Daniel Schmidt, PhD Advisor

March 2021

Acknowledgements

First, I would like to thank Daniel Schmidt, my advisor, and the members of the Schmidt lab for invaluable advice and help. Daniel has been a wonderful and supportive advisor who has given me the freedom to pursue my own ideas. Yungui He has worked by my side and supported everything I've done in lab. I wouldn't have been able to complete my projects without his constant assistance and encouragement.

Wendy Gordon has served as an informal co-advisor throughout my PhD, and provided me with hours of feedback, support, and mentorship. I am thankful for her ongoing support and willingness to open her lab to me without a formal program tie. I am so grateful for the many lunch hours I spent with Eric Aird talking through wild ideas, some of which came to fruition and some of which did not. These talks helped me to grow as a scientist and engineer more than any class I have taken.

I would like to thank my family and friends for their continued support and understanding. My parents have always believed in me and given me the room to grow, without which I would not be where I am today. I'm lucky to have two sisters who are both much smarter than I am, and who inspire me to push myself further every day.

Dedication

To my grandpa, Dr. John P. Zdechlik (piano), who filled my family's life with music.

Abstract

Adeno-associated virus (AAV) is a useful gene delivery tool for clinical and basic research applications. Since its discovery nearly 60 years ago, it has become a popular vector because of its small size and low immunogenicity. However, natural tissue tropisms are limited and frequently not useful. Past work modifying the capsid has been limited by structural constraints and a lack of modularity. I aimed to address this problem by creating a modular retargeting system in which the AAV capsid can be rapidly retargeted to any given antigen without newly mutating capsid proteins. I characterized AAV-antibody composites produced by incorporating a small DNA binding domain into one of the AAV capsid proteins and using chemical conjugation to attach the paired DNA sequence to an antibody. I demonstrated that these antibody-AAV conjugates are capable of infecting cells via the antibody-antigen interaction in immortalized and primary cells. Additionally, I created six capsid variants incorporating small targeting scaffolds into each AAV capsid protein. These variants will enable future researchers to select variants from scaffold libraries *in vivo*, making more specific, better targeting moieties.

Using these retargeted vectors, I worked with colleagues to target prostate tumors *in vivo* and to deliver new payloads with therapeutic potential. A collaborator had recently developed an antibody against a prostate tumor stromal marker, fibroblast activation protein alpha (FAP). By conjugating this antibody to my modified virus, we were able to deliver a fluorescent marker specifically to the

tumor. Next, we plan to deliver therapeutic payloads to treat the tumor with minimal off-target effects. Additionally, I assisted in designing and testing a version of prime editor suitable for delivery by AAV. While the efficiency of this tool is still relatively low, it represents an important starting point for adaption of prime editor for use *in vivo*. In future work, we hope to use antibody-targeted AAV to deliver this and other editing reagents in murine disease models.

Finally, to further investigate the landscape of the AAV capsid, we prepared a domain insertion library using a protocol recently developed by the lab. The results from this study will provide valuable information about the plasticity of the AAV capsid that can be used by future researchers to create new variants with added functionality.

Table of Contents

Acknowledgment……i

Dedication……ii

Abstract……iii

Table of Contents……v

List of Figures……vii

List of Tables……ix

Chapter 1: Introduction……1
Introduction……1
Figures……14

Chapter 2: Programmable assembly of AAV-antibody composites for receptor-mediated gene delivery……17

Preface……18
Introduction……20
Materials and Methods……25
Results……34
Discussion……46
Figures……52

Chapter 3: Split *Staphylococcus aureus* prime editor for AAV delivery……66

Preface……67
Introduction……68
Materials and Methods……72
Results……79
Discussion……87
Figures……90

Chapter 4: *In Vivo* gene delivery to prostate tumors via an anti-fibroblast activation protein antibody-AAV composite……108

Preface……109
Introduction……110
Materials and Methods……114

Results……..……117
Discussion and Future Directions……..119
Figures……..……121

Chapter 5: Oligo Library Synthesis-Mediated Mutational Scanning of Adeno-Associated Viral Capsids…….……125

Preface……....126
Introduction……..……..…..127
Materials and Methods..……..…132
Results, Discussion & Future Directions……..…136
Figures……..……138

Chapter 6: Conclusions and Future Directions…….……141

Bibliography...…….……145

List of Figures

Chapter 1: Introduction

Figure 1.1 An overview of viral retargeting strategies…….....14
Figure 1.2 Timeline of major AAV capsid engineering events…….....15

Chapter 2: Programmable assembly of AAV-antibody composites for receptor-mediated gene delivery

Figure 2.1 Engineering of AAV capsid proteins…….....52
Figure 2.2 Receptor-mediated infection in a synthetic system…….....53
Figure 2.3 Nanobody-AAV are resistant to neutralizing antibodies…….....54
Figure 2.4 Receptor-mediated transduction of a breast cancer model cell line.....55
Figure 2.5 Generalized receptor-mediated infection using antibody-AAV composites…….....56
Figure 2.6 Efficient retargeting of pre-made AAV…….....57
Figure 2.7 Receptor-mediated infection in primary neuron culture…….....58
Supplementary Figure 2.1 Alignment of AAV capsid proteins from different serotypes.....59
Supplementary Figure 2.2 Trans-complementation plasmids for production of AAV with modified capsid proteins.....60
Supplementary Figure 2.3 Western blot of nanobody- and HUH-modified capsids, and affinity pulldown of nanobody- or mMobA-modified virus.....61
Supplementary Figure 2.4 Antibody conjugation with ssDNA oligo…….....63

Chapter 3: Split *Staphylococcus aureus* prime editor for AAV delivery

Figure 3.1 Enabling precise genome modifications using a split *S. aureus* prime editor …….....90
Figure 3.2 SaPE is capable of diverse edit types across various genomic loci …….....92
Figure 3.3 Split NanoLuc substitutes as an effective dimerization domain and visualization tool …….....93
Figure 3.4 Packaging of split SaPE into AAV …….....94
Supplemental Figure 3.1 Conversion of GFP to BFP …….....95
Supplemental Figure 3.2 Optimization of SaPE using GFP(L202S) reporter system.....96
Supplemental Figure 3.3 Prime editing in U2-OS cells…….....97
Supplemental Figure 3.4 AAV-mediated delivery of tdTomato and SaPE…….98

Supplemental Note 3.1 Cloning SaPE pegRNAs using golden gate assembly ……105
Supplemental Note 3.2 Flow cytometry strategy and representative plots.….107

Chapter 4: *In Vivo* gene delivery to prostate tumors via an anti-fibroblast activation protein antibody-AAV composite

Figure 4.1 Experimental timeline for FAP+ prostate tumor targeting....…….121
Figure 4.2 Analysis of mice seven days post-infection....……...……..122
Figure 4.3 Analysis of infected mice over a ten-day time course.....……123
Figure 4.4 Transmission electron microscopy images of antibody-AAV...…….124

Chapter 5: Insertional mutagenesis libraries of adeno-associated virus to interrogate capsid allostery

Figure 5.1 Workflow for creation of an AAV insertional scanning library ……..138
Figure 5.2 Analysis of AAV-DJ insertional library throughout generation …….139
Figure 5.3 Design of a stable cell line for AAV library production without cross-packaging…….140

List of Tables

Chapter 1: Introduction

Table 1.1 Compiled receptors and tissue tropism of naturally occurring AAV serotypes…….....16

Chapter 2: Programmable assembly of AAV-antibody composites for receptor-mediated gene delivery

Supplementary Table 2.1 Typical titers for recombinantly produced AAV.....….64
Supplementary Table 2.2 Commercial antibodies used in this study ……..65

Chapter 3: Split *Staphylococcus aureus* prime editor for AAV delivery

Supplemental Table 3.1 pegRNA sequences....……...……99
Supplemental Table 3.2 Primers used for genomic DNA amplification.....…….101
Supplemental Table 3.3 SaPE Protein Sequences……..102

Chapter 1

Introduction

Engineering adeno-associated virus for gene delivery

After decades of basic research characterizing molecular pathologies of every disease from sickle cell anemia to glioblastoma, medicine is entering a new era of targeted therapy. With knowledge of the genetic basis of a disorder, it is possible to precisely correct the root cause, rather than just treat the symptoms. “Gene therapy” refers to delivery of a corrected copy of a diseased gene to an affected tissue[1]. There are many ways that an exogenous gene can be delivered to a tissue, broadly falling into viral and non-viral methods. While generally viewed as less immunogenic, non-viral delivery vehicles tend to be less efficient[2]. Adeno-associated virus (AAV) has been the most clinically successful viral gene delivery vector. Two FDA-approved AAV-based gene therapy drugs are currently on the market, and many more are in phase II & III clinical trials[1,3,4]. AAV is a popular choice for gene therapy delivery because of its small size, low immunogenicity, and malleable genome [5]. Since its discovery in the 1960s, AAV has grown from a contaminant to a valuable tool for therapy and research[6]. Decades of research have gone into characterizing its basic biology, engineering its protein capsid, and utilizing its genome to deliver a wide range of tools.

Basic biology of AAV

AAV is a non-enveloped virus with a 25nm diameter protein capsid[7]. The viral genome is 4.7kb single-stranded DNA (ssDNA) capped by hairpin-like inverted

terminal repeats (ITRs). There are three open reading frames encoding a total of eight proteins: four replication, three capsid, and one capsid assembly-activating protein. Although AAV has been studied for nearly 60 years, it wasn't until 2010 that capsid assembly-activating protein was discovered[8]. The two replication proteins, Rep68 and Rep78, are responsible for genome encapsulation and site-specific genome integration on chromosome 19[9–11]. Integration is key for AAV because it must remain latent until the host cells is infected with a helper virus, at which time AAV can replicate[11]. The AAV integration site (19q13.3-qter, *AAVS1*) is in the first exon of myosin binding subunit 5 of protein phosphatase 1[12]. Integration rate is dependent upon multiplicity of infection (MOI), and does not occur consistently enough *in vivo* to be considered a sustainable integration method for engineered payloads[12]. *AAVS1* has been used for integration of many other payloads, as it is known to be a safe harbor[13].

The 60-subunit icosahedral capsid is formed from three proteins encoded by the *Cap* open reading frame. The three capsid proteins, called viral protein (VP) one, two, and three make up the virus in a ratio of 1:1:10, with VP3 composing the bulk of the capsid[14,15]. It is possible for capsids composed solely of VP3 to form, but they are non-infective due to loss of a phospholipase A2 domain encoded in the N-terminus of VP1[16,17]. VP2, however, has been found completely unnecessary for formation of infectious particles[17]. This characteristic makes it a popular target for modification, as disruption of the structure is unlikely to make viral particles less infective and/or stable.

AAV is classified as a dependovirus because it requires co-infection with another virus, typically adenovirus or herpesvirus, to reproduce[7]. Co-infection of cells with two viruses is non-ideal for efficient recombinant production of AAV in research laboratories; the two can be difficult to separate, resulting in adenovirus-contaminated stocks of AAV. To remedy this issue, the minimal components of adenovirus necessary for high-titer AAV production have been identified, and can be expressed on a transfected plasmid or by the virus-packaging cell. In current "helper-free" production of AAV, two adenoviral genes, E1a and E1b, are stably expressed by the packaging cell line, and an additional three necessary genes are transfected on a helper plasmid[18,19].

Structural characterization of AAV capsid

Through x-ray crystallography, the structure of VP3 has been elucidated for many serotypes of AAV[20–23]. A model of the fully assembled capsid can be created from the monomer's structure; recently, cryo-EM data has enabled direct imaging of the full capsid[24]. Some features are shared between serotypes; all have 2-, 3-, and 5-fold axes of symmetry with similar properties and dynamics[25]. Thermal stability and protease sensitivity vary wildly between capsids; it is unclear what functional or evolutionary implication this has[26].

When the virus is formed, the protein capsid assembles before the viral genome is packaged[27,28]. The precise mechanism of this process remains unknown, but it is thought to be mediated by the Rep protein recognizing a region

of the ssDNA genome called the D-sequence. Each genome has two D-sequences, composed of the 20 innermost nucleotides of each ITR that remain linear when the remaining ITRs form T-shaped hairpins[28,29]. Much like the protein capsid, the AAV genome is incredibly information dense. In recombinant production, the entire genome apart from ITRs can be replaced with virtually any sequence of 4.7 kilobases. This size constraint can be limiting in clinical applications, resulting in studies to expand packaging capacity without hindering infectivity that have yielded minimal mixed success[30–33].

The different tropisms of AAV serotypes arise from the viral particles recognizing different cell surface molecules (**Table 1.1**). AAV-2, the most commonly clinically used vector primarily recognizes heparin sulfate proteoglycans via its heparin binding domain (HBD) present in a surface loop of all three capsid proteins[34]. AAV-2 also binds the secondary receptors beta-integrin and fibroblast growth factor receptor. Upon surface binding, the virus is endocytosed in a clathrin-coated pit[35]. After entering the cell, the pH of the endocytic vesicle drops significantly, which causes the N-terminus of VP1 to be externalized. The phospholipase A2 domain breaks apart the vesicle, releasing the viral capsid[36]. It is still unclear whether the viral DNA reaches the nucleus before or after capsid disassembly[37]. Second-strand synthesis of the viral genome must occur before expression and replication[38].

Engineering capsid proteins

While many of its characteristics poise AAV to be an ideal gene delivery vector, the limited number of naturally occurring serotypes leave much room for improvement. The majority of AAV engineering has been focused on making viruses that infect different tissues than natural serotypes and reducing recognition by preexisting serum antibodies[1,39–41]. Increased tissue specificity for virtually any gene improves safety, with lower risk of off-target effects. Specificity is paramount for delivery of immunogenic or potentially harmful gene products such as Cas9 to correct a genomic mutation in a diseased tissue, or delivery of a therapeutic/apoptotic gene to a tumor. Retargeting efforts can be basically characterized into six categories and broadly defined as either structure-guided or library-based (**Figure 1.2**)[42].

Rational AAV capsid engineering

Incorporation of a targeting moiety as a means of facilitating infection can be categorized as direct targeting. The first example of AAV being retargeted to infect a specific tissue dates to 1998, when the single chain variable fragment (scFv) of an anti-CD34 antibody was introduced to the N-terminus of AAV-2 VP2 as a chimeric protein (**Figure 1.2**)[40]. The engineered anti-CD34 AAV was shown to infect CD34+ cells several orders of magnitude more effectively than unmodified AAV2.

Taking this one step further, several groups have developed a series of directly targeted vectors based on designed ankryrin repeat proteins (DARPins) and affibodies. DARPins are a class of 14 kDa single domain proteins that can be selected to bind a target protein. They are advantageous because of their functional simplicity relative to derivatives of naturally occurring targeting molecules (i.e. scFv or nanobodies)[43]. Two anti-EGFR AAV-2 variants were created by incorporation of an anti-EGFR DARPin or affibody into the N-terminus of VP2[41]. Both of these capsids infected cells over-expressing EGFR more effectively than AAV-2. Two engineered capsids containing DARPins targeting Her2 and EpCAM fused to the N-terminus of VP2 were described one year later[44]. These vectors were different in that they also contained a tag that allowed for affinity purification. Again, both effectively infected cells expressing the targeted receptor more effectively than unmodified AAV-2.

The main critique of direct targeting is that it is a resource-intensive, high-risk endeavor[42]. Ultimately, the goal of any retargeted vector is clinical use, and the journey from bench to bedside is arduous. If the engineered capsid behaves unexpectedly at any point in pre-clinical or clinical trials, little can be done to modify it without starting the whole process over. For this reason, high-throughput, data-driven engineering approaches have gained popularity in the last decade.

Library-based AAV engineering

While structure-guided, rational design can be a powerful method of altering viral capsid properties, library-based, structure-blind approaches have been used to create many new and differently functional variants. Some of the earliest AAV libraries were created by randomly recombining regions of different AAV serotypes to select for a vector with properties of each parent. An example of a created mosaic capsid is the creation of AAV-DJ, a serotype selected from a library of shuffled AAV-2, 4, 5, 8, and 9 capsids[39]. By applying a selection pressure, a capsid with very specific properties can be generated. AAV-DJ was selected to be minimally immunogenic with high liver tropism, properties that have led to it becoming one of the most commonly used synthetic serotypes. Further engineering has been done to improve its biodistribution[45].

A more modern example of capsid engineering is the creation of AAV-PHP-B by the Gradinaru group[46]. Starting with AAV-9 as a base, a randomized seven amino acid loop (heptamer) was inserted into the surface of the capsid. By extracting AAV genomes from the mouse brain, the group was able to determine variants that could effectively cross the blood-brain-barrier (BBB). Synthetic evolution of a delivery vehicle faces major limitations in interspecies translation. While the AAV-PHP.B capsids work in the mouse line used for their selection (B6), a follow-up study found that they could not cross the BBB in other animal models, or even in other mouse lines[47]. The mechanism of transduction was characterized to be dependent upon Ly6A, a GPI-linked membrane protein which is absent in some mouse lines[48,49]. Notably, there is no known human homologue of *Ly6A*,

making it difficult to believe that AAV-PHP.B would be able to cross the blood-brain-barrier in humans[50]. Given the volume of proteomic data for both model species and humans, it seems appropriate to create new capsid variants with a known mechanism of transduction.

One major barrier to rational engineering of AAV has been the delicate constitution of the protein capsid. One group sought to map permissive spots in the capsid of AAV2 by inserting the coding sequence for mCherry at each position through VP3[51]. Only one loop of the capsid could accommodate mCherry while still forming a functional virus, which is illustrative of the fragility of the capsid. No further domain insertion studies have been completed, either investigating AAV-2 or other serotypes, but we anticipate that a more comprehensive study will be published in the next few years (see chapter 5).

Mutational scanning has also been applied to AAV-2, revealing that even single point mutations can significantly change biodistribution and immunogenicity of the viral capsid[52]. Large, comprehensive screening studies such as this are beneficial in two ways: they provide directly useful structural insights, and they collect a bulk of information that can be further analyzed with different goals in mind. Recently, a massive library with a diversified 28 amino acid region of AAV-2 was characterized, and data used to predict capsid viability using machine learning[53]. While there is no directly useful piece of information resulting from that work, the idea of first diversifying and screening variants *in silico* represents a

powerful new tool. Other groups have applied machine learning to predict capsid properties from massive data sets[54].

Clinical gene therapy applications of AAV

Beyond the bench, AAV is one of the most promising methods of clinical gene delivery. There are currently 145 ongoing clinical trials using AAV as a gene delivery vehicle registered with clinicaltrials.gov, five of which are treating cancer (March 11 2021). In late 2017, the FDA approved the first AAV-mediated therapeutic, a treatment for Leber congenital ameruosis type 2[55]. This drug is an AAV2 capsid carrying the *RPE65* gene and is administered as a sub-retinal injection.

In the European Union, alipogene tiparvovec (Glybera), a mutated form of lipoprotein lipase delivered by AAV1, was approved for clinical use in 2012[56]. Lipoprotein lipase deficiency (LPLD) is a disease characterized by high concentrations of fat droplets in the blood, which can lead to pancreatitis and development of type II diabetes. The drug was administered by multiple intramuscular injections and caused minimal side effects. Circulating lipids were significantly reduced in half of the patients two weeks post-injection, however in most patients, levels returned to pre-treatment within three months. This limited success, small patient population, and high cost led to the treatment being discontinued in 2018[57].

The most recently approved AAV-based therapeutic is onasemnogene abeparvovec (Zolgensma) to treat spinal muscular atrophy (SMA). SMA is caused by mutations of the gene *survival motor neuron 1* (*SMN1*) and leads to death or permanent ventilation within the first 14 months of life[58]. Zolgensma is an AAV9 capsid encoding *hSMN1* administered intravenously as one dose. Early results have shown striking improvement in symptoms, including an increase in probability of event-free survival from 20% in untreated to 100% in treated patients in a phase I trial[58]. Zolgensma is currently the most expensive medicine ever approved by the FDA at around $2 million per treatment spread out over 5 years [3]. Gene therapies pose a unique challenge to pharmaceutical price models as one-and-done "cures". Prior to its release, cost-effectiveness analysis of Zolgensma compared to standard of care treatment revealed that it would have been a cost-effective replacement if priced at $5 million or less[4]. While this garnered significant media attention as a high price for a single treatment, the lifetime benefit to patients and their families is undeniable. Gene therapy is ushering in a new era of pharmaceutical business models that will be fascinating to watch unfold.

AAV as a delivery vector for editing reagents

In an interview months before receiving the Nobel Prize for her discovery and characterization of CRISPR/Cas9, Jennifer Doudna cited the two primary challenges to the technology were delivery and DNA repair[59]. For safe *in vivo* use, it will be necessary to deliver editing reagents specifically to only cells requiring

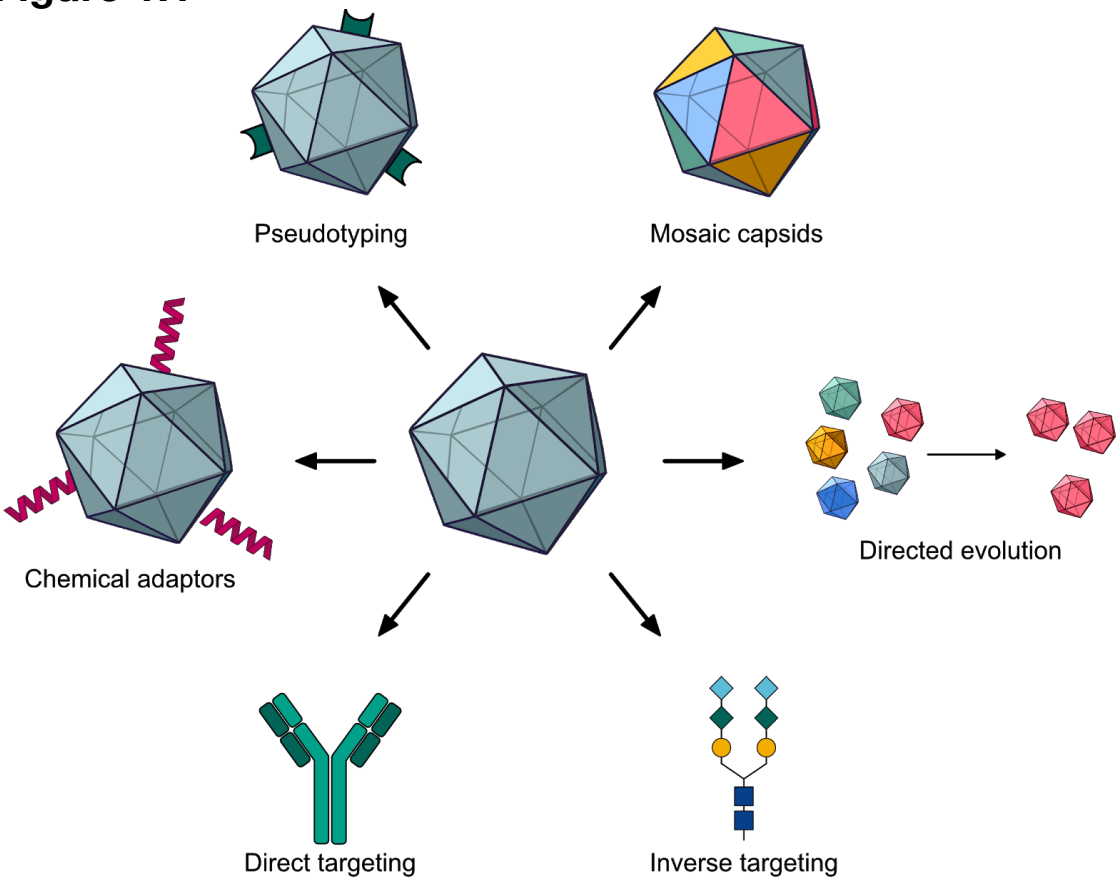
editing. AAV is a top contender for delivery of editing reagents for the same reasons it is desirable for gene therapy[2]. Many editing therapeutics currently in clinical trials utilize *ex vivo* editing, where cells of interest are transduced with retroviral vectors before being returned to the patient[60]. The long-term safety of cells transduced with retroviruses is somewhat unknown. Recently, a clinical trial using a retrovirus to deliver editing reagents to patient-derived CD34+ hematopoietic stem cells for treatment of sickle cell disease was halted after two of the 14 patients treated developed leukemia[61]. Whether the retrovirus or some other factor caused the patients to develop cancer is unclear, but it raises serious concerns regardless.

One of the main drawbacks of AAV for delivery of editing reagents is its genome size. Neither full-length *S. pyogenes* Cas9 nor APOEC-derived base editors fit in the genome with expression elements. Split versions of Cas9 and base editors have been developed to circumvent this issue[62–64]. Additionally, a smaller version of Cas9 from *staphylococcus aureus* has been characterized specifically for use with AAV[65]. Using these modified editors, high levels of phenotypic and genotypic modifications have been achieved *in vivo*[64,66,67]. Splitting cargo between two vectors can allow for greater control of expression, either by combinatorial targeting or by temporally regulating when each component is delivered[68]. While no AAV-based editing therapeutics are currently in clinical trials, that will likely change within the next decade[2].

Viral gene therapy for treatment of cancers

The idea of using infection to treat cancer— whether through directly infecting malignant cells or more broadly engaging and recruiting the immune system— is far from new, dating back to the mid-1800s[69]. Modern viral therapies for cancer broadly fall into three camps- using the virus' natural lytic infection cycle to cause cell death, delivering an engineered lethal payload using the virus as a delivery vector, or recruiting the immune system[70]. Viral cancer gene therapy, delivering an engineered payload to tumor cells using a viral vector, is a promising treatment option that has been explored for multiple tumor types, including prostate cancer, glioma, and liver cancer[71–73]. The payload delivered varies, and is frequently either a suicide gene like HSV thymidine kinase (HSV-TK) or an immunostimulatory gene[74]. Different viruses have different utility in cancer therapy, with some more effectively recruiting the immune system, and others more effectively delivering therapeutic payloads. As with any gene therapy, one of the main hurdles is specific delivery to the tissue of interest. To address this problem, researchers have introduced regions that can be recognized by cancer-associated membrane markers such as integrins or CD13 into the viral capsid[75,76]. While effective, these markers are not applicable to all tumor types, so there is much work yet to be done to target cancer gene delivery vectors.

Figure 1.1



Broad categories of AAV engineering. Adapted from[42] AAV capsid engineering efforts can largely be categorized as structure-guided or library-based. *Structure guided*: Pseudotyping is replacing a capsid protein with one from a different serotype. Chemical adaptors can be added to an unmodified capsid post-production in cells. Direct targeting attaches targeting moieties (i.e. an antibody, scFv, nanobody, DARPin, etc.) to the capsid. Inverse targeting *Library-based*: Mosaic capsids are selected from libraries composed of shuffled regions of multiple serotypes. Directed evolution iteratively selects from a library of capsid variants with diversity typically introduced randomly at the DNA level.

Figure 1.2

Timeline of AAV engineering milestones since 1998. [40,41,46,77,78]

Table 1.1

Serotype	Primary receptor	*In vivo* tissue tropism
AAV1	α2-3 and α2-6 N-linked sialic acid (SIA) [23,79]	Liver, muscle, airway epithelial cells, hematopoietic stem cells, CNS, pancreas [23,80–82]
AAV2	Heparin sulfate proteoglycan (HSPG), FGFR1, αVß5, laminin receptor (LamR)[79,83,84]	CNS (limited), retina, muscle, lung, liver, hematopoietic stem cells, carotid arteries, intestinal epithelial cells, pancreatic beta cells, salivary glands, kidney [79,85]
AAV3	HGFR, HSPG[79,86]	Xenografted hepatocellular tumors, retina, lung, liver (limited) [82,86,87]
AAV4	α2-3 O-linked SIA [79]	Kidney, brain, muscle, heart, lung [85]
AAV5	α2-6 N-linked SIA, PDGFR [79,80]	Muscle, CNS, hematopoietic stem cells, retina, pancreas [80,82]
AAV6	α2-3 and α2-6 N-linked SIA, HSPG, EGFR [23,79]	Liver, muscle airway epithelial cells, CNS, heart [23,82,85]
AAV7	Unidentified [79,88]	Liver, hind-limb skeletal muscle, testes [82,85]
AAV8	LamR[79]	Liver, muscle, CNS, retina, pancreas [82,89]
AAV9	N-linked galactose of SIA, integrins [79]	Heart, lung, testes, liver, hind-limb skeletal muscle, CNS [82,85]

Compiled information about each major natural AAV serotype

Chapter 2

Programmable assembly of AAV-antibody composites for receptor-mediated gene delivery

Alina C. Zdechlik[1], Yungui He[2], Eric J. Aird[1], Wendy R. Gordon[1] & Daniel Schmidt[2]

[1]Dept. of Biochemistry, Molecular Biology & Biophysics, [2]Dept. of Genetics, Cell Biology & Development, University of Minnesota, Minneapolis, MN 55455

ACS Bioconjugate Chemistry. PMC7676631
DOI: 10.1021/acs.bioconjchem.9b00790 **Selected for the issue's cover image**

Author Contributions
A.Z. and D.S. designed the study. W.R.G advised on HUH technology. A.Z. conducted the experiments with Y.H.'s assistance. E.J.A and W.R.G. consulted on implementation of gene editing experiments. A.Z. and D.S. analyzed the data and authored the manuscript. All authors have given approval to the final version of the manuscript.

Preface

Adeno-associated virus (AAV) has emerged as a viral gene delivery vector that is safe in humans, able to infect both dividing and arrested cells, and drive long-term expression (> 6 months). Unfortunately, the naturally evolved properties of many AAV serotypes– including low cell type specificity and largely overlapping tropism – are mismatched to applications that require cell type-specific infection, such as neural circuit mapping or precision gene therapy. A variety of approaches to redirect AAV tropism exist, but there is still need for a universal solution for directing AAV tropism toward user-defined cellular receptors that does not require extensive case-by-case optimization and works with readily available components. Here, we report AAV engineering approaches that enable programmable receptor-mediated gene delivery. First, we genetically encode small targeting scaffolds into a variable region of an AAV capsid, show that this redirects tropism toward the receptor recognized by these targeting scaffolds, and also renders this AAV variant resistant to neutralizing antibodies present in non-human primate serum. We then simplify retargeting of tropism by engineering the same variable loop to encode a HUH tag, which forms a covalent bond to single-stranded DNA oligos conjugated to store-bought antibodies. We demonstrate that retargeting this HUH-AAVs toward different receptors is as simple as 'arming' a pre-made non-infective AAV template with a different antibody in a conjugation process that uses widely available reagents and requires no optimization or extensive purification.

Composite antibody-AAV nanoparticles structurally separate tropism and payload encapsulation, allowing each to be engineered independently.

Introduction

Adeno-associated viruses (AAV) are members of the Parvoviridae family of ssDNA viruses that can infect both dividing and non-dividing cells[90]. AAVs are not associated with disease in humans[9], and the general safety of AAVs for human gene therapy has been established in numerous clinical trials [91]. AAVs are also well tolerated by laboratory animals [92]. Together, these factors have contributed to the widespread use of AAVs in clinical and research applications, for example investigating and treating neurological, neurodegenerative, and neuropsychiatric disorders [93,94]. Several natural AAV serotypes exist, which differ in viral capsid sequence [95]. These sequence differences manifest as different properties of AAVs' 20-25 nm diameter icosahedral virion capsid, and much of the application-relevant biology of AAVs is derived from these properties. This includes capsid/receptor interactions, which drive tissue tropism[96,97], and antigenicity and pre-existing immunity in mammals[98,99]. Cell surface glycans have been identified as the primary cell attachment receptor for most naturally occurring AAV serotypes (reviewed in [97,100]), and the recently identified AAV receptor (AAVR) is critical for endosomal entry[35]. Both are present on many cell types and therefore do not provide specificity of viral transduction beyond tissue specificity of different serotypes. The evolved properties of natural AAV serotypes are therefore a practical limitation for AAV-based gene delivery when cell type specificity is paramount, and when ectopic expression in off-target cell types is an undesirable safety risk[7].

Pioneering studies have tackled this problem in several ways, all made possible by recombinant production methods for AAVs that only require a genetic payload flanked by two inverted terminal repeats (ITR) as the *cis*-element[101]. All other elements, including the *cap* gene, which encodes the three viral AAV capsid proteins VP1–3, can be provided in *trans* (Figure 1a). 60 copies of these beta-barrel capsid proteins at a ratio of 1:1:18 form an icosahedral virus shell [30]. VP1 is essential for viral particle infectivity[16,102], VP2 is a redundant and nonessential protein[17], and VP3 is the most abundant capsid protein. One line of research has attempted to change AAV serotype tropism by resurrecting ancient serotypes [103] or repeated mutagenesis and shuffling of highly similar capsid genes[21]. One capsid gene shuffling experiment, using AAV-2, -8, and -9 as input, resulted in AAV-DJ, which shows superior infectivity in culture and biodistribution predominantly limited to the liver[39].

Another approach focused on the addition of short peptides in various capsid regions. From scanning mutagenesis studies[25,51] and analysis of crystal structures of AAV capsids (e.g.[20,104]) a map of variable regions (VR) has emerged (reviewed in[105]). Best known is VR VIII (Figure 1b, Figure S1) a region that contains a heparin binding domain (HBD) critical for the heparin sulfate directed tropism of AAV2. Insertion of a 14 amino acid integrin-targeting peptide, L14, was among the first demonstrations that capsid tropism could be rationally retargeted toward a user-specific cell type[104]. Inserting random peptides into this loop to create a library of AAV variants is called viral display, and this method can be used to select AAV

capsid modifications that redirect AAV tropism toward, for example, cancer cells and endothelial cells[106,107]. Viral display was recently combined with *in vivo* biopanning into a more efficient directed evolution approach to select from a randomized AAV capsid library variants with specific tissue tropism[108]. Another viral display study created AAV-PHP.B, which efficiently crosses the blood brain barrier after intravenous injection[46] in some mice strains[109–111] and rats[111,112]. The inability of AAV-PHP.B to cross the blood-brain-barrier in genetic backgrounds of mice that lack the GPI-anchored cell surface protein LY6A[47–49], and the mixed result of using AAV-PHP.B in non-human primates (severe toxicity at high doses[47,113], lack of CNS tropism[114]) are cautionary notes. They highlight the complex and often unknown link between genetic variation of capsid genes and the selected phenotypes in directed evolution approaches, which may result in inadvertent optimization for undesired traits.

A third approach to redirect AAV tropism is to add larger targeting scaffolds, most often to the N-terminus of VP2. This first worked to include a single chain antibody fragment against CD34[40], and later for adding an CX3CL1 chemokine binding domain, the human hormone leptin, and even GFP (30kDa)[17,115]. The same approach was recently adopted to insert into VP2 a designed ankyrin repeat (DARPin)[44,116,117] or an Affibody[41], which directed virus tropism toward human endothelial growth factor receptors expressing cells. Unfortunately, modification on VP2's N-terminus usually lowers the efficiency of VP2 incorporation into the capsid, increases the fraction of capsid that are empty (i.e. do not contain a genetic

payload), and the modified N-terminus can be trapped inside the viral capsid[17,25,40]. A variation on the theme of introducing specific targeting scaffolds into AAV capsids is the addition of an *S. aureus* protein A derived immunoglobulin binding domain (Z34C) into VR VIII[77]. This allowed subsequent coupling to antibodies against β1 integrin, CD117, or CXCR4, and redirected AAV tropism toward cell lines expressing these cell surface receptors. It is worth noting that the interaction between Z34C is a non-covalent, albeit high affinity, interaction. The dissociation rate constant is sufficiently high (~$3x10^{-2}$ s^{-1})[118] that in an *in vivo* setting, it is a real possibility that Z34C-AAV-coupled antibodies are swapped out for endogenous antibody, which could lead to significant loss of infectivity or off-target infection. More recently, the addition of a split-intein domain the N-terminus of VP2 allowed covalent linkage of targeting ligands (including single chain antibodies (scFVs) and DARPins) to AAV[119]. However, this requires production and purification of targeting ligand/split-intein fusion proteins.

Thus, while a variety of approaches to redirect AAV tropism exist, many require the repeated engineering capsid proteins (viral display, targeting scaffold insertions). Those that take a platform approach, form non-covalent adducts that limit biologic stability (e.g., AAVs displaying antibody targeting scaffolds) or require custom-made targeting ligands (e.g., intein-scFV fusions). There is still need for a universal solution for directing AAV tropism toward user-defined cellular receptors, and in turn cell types and tissues, that does not require extensive case-by-case optimization and works with readily available components.

In this study, we combine the strategies of previous studies to simplify retargeting of an engineered AAV variant. After validating an understudied AAV surface loop for insertion of small targeting scaffolds, we introduce into the same loop of the viral capsid a HUH tag. HUH tags are domains that form covalent bonds to single-stranded (ss) DNA in a sequence specific fashion[120]. If these ssDNAs are conjugated to antibodies, we can form covalent antibody-AAV composites whose tropism is determined by the antigen-specificity of the linked antibody. We demonstrate that retargeting antibody-AAV composites toward different receptors is as simple as 'arming' a pre-made AAV template with different store-bought antibodies in a 30-minute reaction that requires no optimization or extensive purification. Composite AAV nanoparticles structurally separate tropism and payload encapsulation, allowing each to be engineered independently.

Methods & Materials

Construction of separate VP expression plasmids:

All primers were designed using NEBaseChanger or NEBuilder. The CDS of all plasmids were verified by Sanger sequencing. Backbone integrity of all DNA preparations was confirmed by analytical digest.

pCMV-cap was created by subcloning the *cap* open reading frame of AAV-DJ into a pcDNA5 backbone. The heparin binding domain (amino acids Arg587-Arg590) was mutated to an HA tag (amino acid sequence YPYDVPDY) or a 6xHIS tag using NEB Q5 site-directed mutagenesis. To create plasmids individually expressing each VP, the start codons of the other two VPs were individually mutated using NEB Q5 mutagenesis: Met1Lys to knock out expression of VP1, Thr138Ala to knock out expression of VP2, and Met203Lys, Met211Lys, and Met235Lys to knock out expression of VP3. We found that it was necessary to mutagenize all three VP3 start codons to fully ablate expression. Using these mutations, we created staging plasmid pCMV-VP1ΔHBD>HA/6xHIS, pCMV-VP2ΔHBD>HA/6xHIS, and pCMV-VP3ΔHBD>HA/6xHIS. Coding sequences for an anti-GFP nanobody (gift of Sivaraj Sivaramakrishnan), Gp2[58], and mMobA[49] were inserted, adding a $(GGGGS)_2$ flexible linker on either side, into staging plasmids for each VP separately using NEB HiFi assembly.

Production of modified viruses:

AAV was packaged using established a helper-free production protocol by the UMN Viral Core. Briefly, AAV293 cells at 60% confluence were transfected with 600 µg of DNA (viral shuttle vector encoding the payload, helper plasmid, rep/cap plasmids at 1:1:1 ratio) using polyethyleneimine. 24 hours after transfection, media was changed and cells were checked for fluorescent protein expression (when applicable) to confirm transfection. 72 hours after transfection, cells were detached and pelleted. Viral particles were released from producer cells by repeated freeze/thaw cycles in the presence of Benzonase (100 units). Crude lysates were cleared by centrifugation and further purified using sucrose gradients.

Viral particles in the supernatant were titered using qPCR with WPRE-, GFP-, or tdTomato- specific primers. Kanamycin-specific primers were used to confirm absence of plasmid DNA after Benzonase treatment.

The pAAV-CAG::tdTomato (codon diversified) shuttle vector was a gift from Edward Boyden (Addgene plasmid # 59462). The pAAV-CAG::GFP shuttle vector was a gift from Edward Boyden (Addgene plasmid # 37825).

Western Blotting:

$2x10^{11}$ genome copies (g.c.) of each virus was denatured by heating to 95 C for 5 minutes and run on SDS-PAGE. For paired antibody blots, the same reaction was split between two wells to yield an identically treated replicates for each antibody (B1 or HA.11). After staining with primary antibody (B1 or HA.11, see

Supplemental Table 2.2), blots were treated with anti-mouse horseradish-peroxidase coupled secondary antibody, and visualized using ECL reagent. Blots were imaged using an Amersham Imager UV600.

Virus Pulldown assays:

For pulldown with streptavidin as the affinity matrix, approximately 1x10 g.c. of VP1- or VP2-mMobA-AAV was reacted with 1 nmole of biotinylated mMobA-specific ssDNA in HUH reaction buffer (1:20 salmon sperm DNA, 1 mM $MgCl_2$) for 30 minutes at 37 degC. 50 µL of magnetic streptavidin beads were pelleted using a strong magnet and washed three times with 300 µL wash/binding buffer (0.5 M NaCl, 1 mM EDTA, 20 mM TrisHCl pH 7.5) to remove storage buffer. Oligo-conjugated virus was bound to streptavidin beads for 10 minutes at RT with gentle agitation. Beads were washed three times with 300 µL wash/binding buffer. Beads were suspended in 100 µL wash/binding buffer with virus still bound and treated with proteinase K to extract viral genomes, as it is was not feasible to elute from the streptavidin. Beads were pelleted and supernatant was processed using Zymo Clean and Concentrator 5 kit prior to qPCR. qPCR was used to measure the number of viral genome copies retained by streptavidin beads.

For pulldown with GFP as the affinity matrix, 150 µL of of magnetic streptavidin beads were pelleted using a strong magnet and washed three times with 300 µL of wash/binding buffer. 800 pmoles of a biotinylated superfolder GFP (recombinantly produced) were bound to washed beads at room temperature. The

beads were then washed with wash/binding buffer to remove unbound GFP. Separately, approximately $1x10^{10}$ g.c. of a VP2-mMobA virus was reacted with an anti-GFP antibody as previously described. For nanobody-AAV GFP pulldowns, virus was not treated prior to binding to beads. Anti-GFP virus was then incubated with GFP-streptavidin beads for 10 minutes at room temperature with gentle mixing. Beads were washed twice with wash/binding buffer and resuspended in 100 µL buffer for treatment with proteinase K and benzonase to extract vector genomes for qPCR.

Nanobody infection of HEK293-FT:

HEK293-FT cells were plated on growth factor-depleted matrigel in a 24-well cell culture plate and reverse transfected with pCAG-GFP-GPI[53] or a control plasmid (pATT, a gift from David Savage, Addgene plasmid # 79770). 24 hours after transfection, media was removed and cells were washed with 500 µL DMEM. Cells were incubated with DMEM and virus at a total volume of 240 µL for one hour at 37 degC, after which 1 mL D10 cell maintenance media (DMEM, 10% w/v fetal bovine serum (FBS), 1% w/v sodium pyruvate, 1% w/v penicillin/streptomycin) was added on top of DMEM/virus. 48 hours after infection, cells were analyzed by fluorescent microscopy and flow cytometry. For neutralization assay, viruses were incubated in media with a given percentage of serum at 37 degC for 30 minutes before addition to cells. Culture medium was removed and cells were washed with DMEM before addition of serum/virus. 24 hours after infection, virus/serum was

replaced with unmodified D10. 48 hours after infection, cells were analyzed by fluorescent microscopy and flow cytometry.

Gp2 infection of HEK293 expressing hIR:

HEK293-hIR/GFP and HEK293-eGFP cells were a gift from Douglas Yee [58]. Each cell line was plated on growth factor-depleted matrigel in a 24-well plate. 24 hours after plating, media was removed and cells were washed with 500 µl DMEM. Cells were incubated with DMEM and virus (VP2-Gp2IR-AAV) at a total volume of 240 µl for one hour at 37 C, after which 1 mL D10 was added on top of DMEM/virus. 48 hours after infection, cells were analyzed by fluorescent microscopy and flow cytometry.

Conjugation of HUH-AAV to antibodies:

Antibodies (**Supplemental Table 2.2**) were conjugated to HPLC-purified ssDNA (CCA GTT TCT CGA AGA GAA ACC GGT AAG TGC ACC CTC CCT GAT GA - AmMO) using the Innova Biosciences Thunderlink Plus kit according to manufacturer instructions. Antibody and DNA were combined to an anticipated ratio of 3:1 DNA:antibody. Conjugation efficiency was determined via reducing SDS-PAGE. Average ratio of oligo:antibody was calculated by quantifying intensity of each band using ImageJ (release 1.52r). For GFP antibody, experimental DNA:antibody ratio was 2.29, for EGFR antibody 1.83, for CD7 1.40. It was not determined for L1CAM.

AAV Composite assembly:

HUH-AAV in PBS with 10% glycerol was incubated in 1:20 salmon sperm DNA, 0.2 M $MgCl_2$, and 10 pmoles antibody-oligo for 30 minutes at 37 degC. Antibody-AAV was applied to cells immediately after composite formation.

GFP antibody-AAV infection of HEK293-FT:

HEK293-FT cells were plated and transfected to express GFP-GPI as for nanobody infection assay. 24 hours after transfection, media was replaced with 200 µl D10. Conjugated anti-GFP-AAV or control conjugated virus was applied to each well in 200 µl D10. 24 hours after infection, 1 ml D10 was added on to each well. 48 hours after infection, cells were analyzed by fluorescent microscopy and flow cytometry.

EGFR antibody-AAV infection of U251 MG:

U-251 MG cells were plated at 20% confluence on growth factor-depleted matrigel in a 24-well plate. 24 hours after plating, cells were treated with conjugated anti-EGFR-AAV or control conjugated virus in 200 µl D10. 48 hours after infection, cells were analyzed by fluorescent microscopy and flow cytometry.

CD7 antibody-AAV infection of Jurkat:

Jurkat cells were cultured in RPMI-1480 with 10% w/v FBS and 1% w/v penicillin/streptomycin for several passages after thawing to acclimate cells to antibiotic. For infection assay, $1x10^5$ cells were seeded into each well of a 96-well treated culture plate in 200 µl media. Anti-CD7-AAV was added to each well. 24 hours after infection, 100 µl media was added and cells were mixed. 48 hours after infection, cells were moved to a 24-well plate and 200 µl media was added. 72 hours after infection, cells were analyzed for transgene and CD7 expression using flow cytometry.

L1CAM antibody-AAV infection of Primary Hippocampal Neuron/Glia Co-Culture

All animal procedures were in accordance with the National Institute of Health Guide for the care and use of laboratory animals and approved by the University of Minnesota Institutional Animal Care and Use Committee (Protocol #1803-35718A). Hippocampal regions from CD-1-022 mice (Charles River Laboratory) postnatal day 0 – 1 were isolated and digested with papain (100 units in Hanks balanced salt solution supplemented with 35 mM glucose, 1 mM Kynurenic acid, 0.3 mg/ml L-Cysteine and 10 mM $MgCl_2$) for 6 – 8 minutes. Cell suspension was washed with Ovomucoid trypsin inhibitor (10 mg/ml), washed three times with 1 mL of plating media (MEM, 10% w/v fetal bovine serum, 0.5% w/v glucose, 10 mM HEPES, 2 mM L-glutamine, 0.5 mg/ml holo-transferrin, 25 µg/ml insulin, B27 supplement, buffered to pH 7.4 with NaOH). The tissue was then mechanically

dissociated by triturating through P1000 plastic pipette tips, and settled by gravitation. Dissociated neurons and glia in the supernatant were plated on Matrigel-coated 24 well glass bottom plates at approximately 50,000 cells per well and maintained in plating medium. Two days after plating, anti-L1CAM-AAV was added to each well.

Immunohistochemistry

Five days after infection, cells were fixed with 2% formaldehyde, washed with PBS, and permeabilized with 1% Triton-100 in PBS at room temperature. Fixed neurons and glia co-cultures were incubated in blocking buffer (PBS, 0.1% Triton-100, 15% FBS) for 30 minutes at room temperature. Fixed cells were stained with primary antibodies (GFAP, clone 2.2B10; NeuN, clone A-EPR12763, see **Supplemental Table 2.2**), diluted to 1:500 in blocking buffer, for 1 hour at room temperature, and washed twice with 0.1% Triton-100 in 2x PBS. Secondary antibodies (GFAP, anti-rat Alexa Fluor 350; NeuN, anti-rabbit Alexa Fluor 700, see **Supplemental Table 2.2**) 1:500 diluted in blocking were added and incubated for 1 hour at room temperature. After washing twice with 0.1% Triton-100 in 1xPBS and twice with 1xPBS, fixed cells were imaged immediately.

Wide-field Fluorescent microscopy:

For unfixed cells, 48 hours after infection, media was removed and replaced with 200 µL Tyrode buffer (125mM NaCl, 2mM KCl, 3mM $CaCl_2$, 1mM $MgCl_2$, 10mM

HEPES, 30mM glucose, pH 7.3). Nuclei were stained with NucBlue Live Cell stain (Thermofisher) for 30 minutes at 37 degC before imaging. NucBlue and Alexa Fluor 350 were excited at 395±25 nm, with a 409 nm dichroic, and recorded at 409 nm longpass. GFP was excited at 470±24 nm, 495 dichroic, and recorded at 525±24 nm. tdTomato was excited at 575±25 nm, 596 nm dichroic, and recorded at 609±54 nm. AlexaFluor 700 was excited at 635±18 nm, with a 652 nm dichroic, and recorded at 680±42 nm. Images were processed using ImageJ (release 1.52r). The Cell Counter plugin was used to determine neuron/glia infection ratios.

Flow cytometry analysis:

Adherent cells were detached using Accutase. All cell types (adherent and non-adherent) were then centrifuged at 400 x g for 3 minutes, washed with 1 mL Tyrode buffer once, resuspended in 300 µl FACS buffer (2% of FBS, 0.1% NaN_3, 1xPBS), and passed through a 35 µm nylon filter tubes.

Samples were run on a BD Fortessa X-20 flow cytometer. GFP was excited with a 488nm laser and recorded with on a 525±50 nm filter. tdTomato was excited with a 561nm laser and recorded on a 585±15 nm filter. APC was excited with a 640nm laser and recorded on 670±30 nm filter. Events were gated on side-scattering area and forward-scattering area and subsequently forward-scattering area and width to separate out live, single cells. Samples were down-sampled to collect 20,000 single cells for fluorescent analysis.

Results

While many studies have focused on engineering VR VIII, VR IV (**Figure 2.1b**) has emerged as an alternative loop amenable to engineering[51,104]. We constructed *cap* variants of AAV-DJ[39], in which VR VIII was mutated to include an HA or 6xHIS tag (AAV-DJΔHBD>HA/6xHIS), which in previous work removed natural tropism of AAV2[41,116,121]. We then introduced an anti-GFP nanobody (nb, size: 13.2 kDa,[122]) into VR IV (amino acid position Thr456). Nanobodies consist of only a single domain that, like antibodies, specifically bind a specific antigen[123]. Since all three AAV capsid proteins (VP1–3) are produced from different open reading frames of the *cap* gene, we generated a set of trans-complementation plasmids with the required mutations of alternate start codons, so we could express each VP independently (**Figure 2.1c, Supplemental Figure 2.2**). Western blot (performed for all VP-nb-AAV) and pulldown assays with GFP-conjugated magnetic beads (VP2-nb-AAV and VP3-nb-AAV only) qualitatively confirmed incorporation of nanobody fusions with all three VPs in rAAV (**Supplemental Figure 2.3a,c**). Nanobody-modified AAVs could be produced to titers that are comparable to wildtype AAV-DJ (~$5x10^{10}$ and ~$4x10^{11}$ genome copies (g.c.)/µl, respectively, **Supplemental Table 2.1**). Note that for VP3-nb-AAV production, we had to co-transfect trans-complementing plasmid encoding for both nb-modified VP3 and unmodified VP3 to achieve acceptable titers. Since VP3 is the most abundant capsid protein (44 copies/capsid), we interpret this as

nanobody insertion having a destabilizing effect that interferes with virus production when capsids are composed of mostly modified capsid proteins.

We next assessed whether nanobody incorporation redirects AAV tropism toward the nanobody's binding target. As a synthetic surface receptor, we expressed GPI-anchored GFP[124] on the extracellular membrane of HEK293 cells and infected these target cells with different virus preparations that all carried tdTomato driven from a strong and broadly active CAG promoter[125] as the genetic payload. Overall, AAV-DJ was more infective at any tested multiplicity of infection (MOI), but nanobody-modified AAV infectivity was within the same order of magnitude (**Figure2.2a-b**). Interestingly, VP2-nb-AAV performed better than VP1- or VP3 modified AAV. We currently do not know what mechanism underlies this difference in infectivity. Possible causes include that a fraction of virions lack nanobody modifications, that nanobody incorporation changes the ratio of empty and full capsid, or that nanobody/receptor interaction direct AAV to a different intracellular compartment after cellular uptake. Arguably, on-target selectivity is the most striking feature of nb-AAV. As expected, when cells where infected with wildtype AAV-DJ, transduction was indiscriminate in both GFP-positive and GFP-negative cells. Conversely, AAV-DJΔHBD>HA proved non-infective, as expected from removing the heparin binding domain that mediates AAV interaction with the cell. AAV capsids with nb-insertion in any capsid protein, on the other hand, preferentially infected GFP-positive cells. As a scoring function for on-target selectivity we calculated the ratio of on-target (GFP-GPI transfected) over off-

target (control plasmid pATT transfected) cells. Selectivity for GFP-positive cells was noticeable already at low MOI (1,000 g.c. / cell) for VP2-nb (score 45.514.68) vs. AAV-DJ (0.810.29). At higher MOI (>1,000 g.c. / cell) VP1-nb and VP3-nb on-target selectivity became apparent, as well. Interestingly, while peak on-target selectivity was similar for all nb-AAV variants, they peaked at different MOI. That selectivity dropped for all nb-AAV at the highest MOI test was driven by higher off-target infection. Overall, these data suggest that VR IV is permissive to nanobody insertion, and that this insertion can redirect tropism toward the receptor target by that nanobody. While overall infectivity it reduced to some degree, high on-target specificity can be achieved at low MOI.

We next tested whether nanobody-mediated targeting toward GFP as a synthetic cell surface receptor provides additional benefits in the forms of evading neutralizing antibodies. That the majority of the human population is sero-positive to most existing AAV serotypes[126] is one of the most pressing concerns with potential AAV gene therapy[7]. Neutralizing antibodies are also an issue when using AAV on non-human primate neuroscience[99]. Neutralization mechanisms of antibodies vary and can include interference with primary receptor binding or post-attachment steps (endosomal escape, conformational change in the capsid, etc.) [80]. When we compared transduction of HEK293 cells expressing cell surface GPI-anchored GFP with either AAV-DJ or VP1-nb-AAV, we found that AAV-DJ transduction was decreased >20 fold in the presence of pooled rhesus macaque

serum (**Figure 2.3**). VP1-nb-AAV transduction, on the other hand, was unaffected even at the highest serum concentration tested (20%). This suggests not only that neutralizing antibodies were present in the rhesus macaque serum, but also that nanobody incorporation provides resistance to neutralization by these antibodies. Interestingly, AAV-DJ is the result of forced evolution in presence of pooled human antisera (purified human Ig (IVIG)), which provides strong negative selection pressure against AAV-2 epitopes, and AAV-DJ transduction efficiency had been found to be unaffected by presence IVIG[39]. We did not test VP2-nb- or VP3-nb-AAV, but suspect that they are resistant to neutralization by serum as well.

We next asked whether this kind of receptor-mediated infection could be extended to other targeting scaffolds. T7 gene 2 protein (Gp2) scaffolds are small (6 kDa) proteins domains with two solvent-exposed loops that can be mutated to generate a binding interface to arbitrary targets[127]. We incorporated into VR IV a Gp2 binding human insulin receptor (hIR[128]) with nanomolar affinity, creating $Gp2^{IR}$-AAV variants. Similar to what we observed for nanobody-AAVs, modification of all three capsid proteins appeared tolerated when judged by virus production titers (7.6–9.6x10^{11} g.c./µl, **Supplemental Table 2.1**). To assess hIR-directed infection, we compared infection of hIR-overexpressing HEK293 and control HEK293 cells that only express a basal (endogenous) level of hIR[128]. While wildtype AAV-DJ indiscriminately infected both hIR-expressing and control cells, and tropism null AAV-DJΔHBD>HA infected neither, all Gp2-displaying AAV at any

MOI preferentially infected hIR-overexpressing cells (**Figure 2.4**). Importantly, when using $Gp2^{IR}$-AAV, we observed greater specificity at lower MOI compared to AAV-DJ; at $1x10^4$ g.c. / cell the ratio of on-target (hIR) to off-target (control HEK293) is >11-fold for VP2-Gp2 (p-value 0.0047, two sided Dunnett's-test for multiple comparison with AAV-DJ as the control). However, efficiency of this transduction was less than 5%. Conversely, at the highest MOI tested, while still significantly more selective than AAV-DJ, the selectivity ratio dropped to ~1.6 (p-value 0.00227, 0.00019, and 0.00073 for VP1-, VP2-, and VP3-Gp2 two sided Dunnett's-test for multiple comparison with AAV-DJ, respectively) while transduction levels approaches 50%. The trade-off between transduction efficiency and on-target selectivity highlights multiplicity of infection is a critical parameter that may require optimization when on-target and off-target cell are defined only by differences in expression levels of the targeted receptor. Overall, our data nevertheless suggest that genetic insertion of targeting scaffolds into VR IV might be a general strategy to redirect AAV tropism. When combined with orthogonal approaches, such as cell-type specific promoters, this may provide an additional axis along which AAVs can be engineered to achieve cell-type specific infection. Our data is also compatible with a simple equilibrium model for virus entry in which infection efficiency depends on expression levels of the targeted surface receptor and the affinity of the encoded targeting scaffold to that the receptor.

While nb and Gp2 insertion into VR IV are well tolerated and redirect AAV infection toward their respective binding target, these require repeated genetic modification and production of viral particles whenever a new targeting scaffold is used. We reasoned that a generalizable method of linking new targeting scaffold to viral particles is achievable by inserting a universal adapter domain into the AAV capsid. Adapter-containing AAV can then be produced once, stored, and 'armed' with different targeting scaffolds as needed. One type of targeting scaffold that is of particular interest are antibodies. Over the past decades research and development in both industry and academic labs have resulted in a rich repertoire of characterized antibodies with well-understood receptor specificity and mature production pipelines [129,130]. Monoclonal antibodies have emerged a class of therapeutic agents to numerous human diseases including cancer, auto-immune, and cardiovascular disorders. And leveraging these existing antibodies for re-targeting AAV tropism might represent a new mode of action for these drugs.

The principle of linking pre-produced AAV to targeting scaffolds has been previously advanced. Early work modifying VR VIII inserted a minimal Protein A Fc-binding domain into the AAV capsid[77]. However, the interaction between this domain and antibodies are non-covalent. Split-intein-mediated covalent attachment of targeting scaffold, on the other hand, requires custom production of components[119].

To establish a method of covalently linking store-bought antibodies to AAV on demand, we turned to the recently described HUH tags, which are small (10-30

kDa) and robustly form covalent bonds with short ssDNA in a sequence-specific fashion[120]. HUH domains are orthogonal and multiplexible; there is minimal overlap between sequences that each HUH domain recognizes and binds. We expected that by displaying a HUH domain on the AAV capsid, we could covalently attach ssDNA-conjugated antibodies to AAV particles in a programmable fashion (**Figure 2.5a**). Modification of antibodies with ssDNA is a widely-adopted step in immuno-PCR and within practical reach to most labs[131].

A set of trans-complementing plasmids allowed us to express individual capsid protein variants with a mMobA HUH tag (size 21 kDa) inserted in VR IV. To determine which HUH insertion in which capsid protein was functional, we reacted AAV particles with mMobA inserted into either VP1 (titer $1.37x10^{11}$ g.c./µl), VP2 (titer $4.9x10^{11}$ g.c./µl), or VP3 (titer $1.5x10^{8}$ g.c./µl) with a ssDNA oligo that contained the mMobA recognition sequences and that was also biotinylated. After pulldown with streptavidin beads, we determined presence of full AAV capsid by qPCR. While we did not recover any fully packaged viral particles when mMobA was inserted into VP3, we did recover particles for insertions into VP1 and VP2 (**Supplemental Figure 2.3c**). In line with our results for nanobody-modified AAV and the lower production titer for VP3-mMobA (**Supplemental Table 2.1**), modification of VP3 appears to be less well tolerated. Alternative explanations for the lack of pulldown are that VP3-mMobA is not conformationally accessible to oligo conjugation (and therefore is not pulled down) or that insertion of mMobA results in mostly empty AAV particles (titers are measures by qPCR with genomic

probes). Our titer data for VP1- and VP2-mMobA showed, however, that insertion of the mMobA HUH tag in the VR IV loop of VP1 or VP2 of AAV-DJ was tolerated. We qualitatively confirmed this for VP2 by Western blot and affinity pulldown (**Supplemental Figure 2.3b-c**). A more quantitative look at the Western Blots suggested that mMobA-modified VP2 is incorporated at a ratio that deviates from the expected 1:1:18 ratio of VP1:VP2:VP3. When using a pan-VP antibody (B1), the band corresponding to VP2-mMobA was very faint, and only with an antibody specific to the HA tag inserted into VR VIII of VP2-mMobA could we confirmed incorporation of the expected fusion capsid protein. Streptavidin-bead pulldown experiment in which the VP2-HUH-AAV was reacted with a mMobA specific oligo that also had was biotinylated similarly suggested that not every particle carried VP2-mMobA at wild type copy number; typically, ~30-40% of virus was retained. Despite the apparently low incorporation of VP2-mMobA, we reasoned that VP1/VP3-only capsid in which the heparin binding domain had been mutated would not contribute significantly to infection and went ahead with functional characterization of VP2-mMobA-AAV (referred to from here on as 'HUH-AAV') without any further enrichment of purification.

To test whether HUH tags allow for on demand assembly of antibody-AAV composites, we conjugated an ssDNA oligo contain the mMobA's recognition sequence to a store-bought GFP antibody (**Supplemental Table 2.2**) using widely available copper-free click chemistry (**Supplemental Figure 2.4**). Composites

between HUH-AAV (carrying CAG-tdTomato as payload) and ssDNA-mAb were formed by brief incubation under physiological conditions. Pulldown experiment with GFP-linked magnetic beads show that GFP antibodies are reacted to HUH-AAV (**Supplemental Figure 2.3c**). We then applied anti-GFP-HUH-AAV composites to HEK293 expressing GFP-GPI without further purification. 48 hours after infection, cells were analyzed for expression of the viral transgene. As anticipated, cells expressing GFP-GPI treated with the full antibody-AAV composite expressed tdTomato, while fluorescence was much lower in cells not expressing GFP-GPI (**Figure 2.5b-c**). Omitting any one component of the composite (i.e. mMobA-specific ssDNA, antibody, GFP-GPI target, mMobA HUH tag) resulted low transduction indicating that only complete composites mediate receptor-specific infection. When compared to nanobody-modified AAV, we did notice slightly higher baseline of off-target infection with VP2-mMobA-AAV (~8% vs. 5% at $1x10^6$ g.c./cell). Since off-target infection was least pronounced in negative controls without VP2-mMobA (i.e., VP1/3ΔHBD>HA, **Figure 2.5c**) we suspect that the biophysical properties of the mMobA domain itself contribute to non-specific infection. That relatively little expression was observed at lower MOI may in part be due to the relative short period of time that was allowed for onset of viral payload expression (48h) and the presence of unreacted free antibody that could interfere with anti-GFP-HUH-AAV attachment to surface-anchored GFP. Most likely, however, relative low infectivity is related to heterogeneity in the composition of the full antibody-AAV composites, which we suspect comes from

multiple sources. For one, VP2-mMobA fusion proteins may not incorporate into assembling AAV particles (as suggested by Western Blot, **Supplemental Figure 2.3b**) or may decrease particle stability (although, we did not see a decrease in production titer, **Supplemental Table 2.1**). In both cases, HUH-AAV ready to react with ssDNA-conjugated AAV represent only a fraction of all AAV particles. HUH incorporation may also prevent encapsulation of the viral genome, in which case HUH-AAV particles are more likely to be empty and, while infective, non-transducing. Again, this is a less likely scenario as we did not observe a decrease in production titer. Finally, mMobA reaction efficiency with its cognate ssDNA target is less than 100% efficient (as suggested by pulldown experiments with biotinylated mMobA-specific ssDNA, **Supplemental Figure 2.3c**). All of these inefficiencies can add up and lead to a mismatch between qPCR-determined titers (which measure full AAV particles and on which we based MOI) and the indeterminate titers of fully-assembled antibody-AAV composites that encapsulate a viral payload. Clearly, process improvements (removing unreacted antibody, enriching for HUH-containing AAV, etc.) could boost infectivity. However, even in light of the current limitations, our data demonstrate that incorporation of HUH tags into AAV capsids allows for covalent attachment of antibodies, and that these antibodies redirect tropism of antibody-AAV composites toward specific cellular receptors.

To demonstrate ease of re-targeting to a new cell surface receptor, we next formed composites with HUH-AAV from the same batch used in the experiments

above, but now linked to an EGFR antibody or a CD7 antibody, both store-bought. Anti-EGFR-AAV composites infectivity was tested in U-251 MG cells, which are an immortalized glioblastoma cancer stem cell line that naturally expresses high levels of EGFR[132] (**Figure 2.6a**). As before, when antibody was absent or present but not covalently linked to the AAV capsid, only minimal infection was observed. The full composite was about half as infectious as AAV-DJ. We also used Jurkat cells, an immortalized T cell line that naturally expresses high levels of CD7, for anti-CD7 AAV composites infection[133] (**Figure 2.6b**). Interestingly, AAV-DJ show only marginal infection (10% at $1x10^7$ g.c./cell) making this a good example for how generalizable receptor-mediated infection with antibody-AAV composites can be. As before, we found that mutations of the heparin binding domain (AAV-DJΔHBD>HA) almost completely abolished infectivity even at high titer. Providing but not covalently linking an anti-CD7 antibody to the HUH-AAV also did not result in infectious particles beyond putative non-specific levels. Full anti-CD7-AAV composites, on the other hand, could achieve >80% infection at the highest MOI tested. Overall, these data suggest that HUH incorporation is a generalizable method to redirect AAV tropism toward any user-selected surface receptor in a programmable fashion.

To assess whether receptor-mediated infection extends beyond synthetic targets and immortalized cell lines that express a cell surface receptor at relatively high levels, we turned to primary neuron hippocampal neuron cultures. Neuron cultures are prepared by enzymatic and physical dissociation of brain tissue from newborn

mice and are actually a co-culture of neurons (~70% of cells) and glia (~30% of cells). To develop a viral targeting strategy that can differentiate between neurons and glia by virtue of a cell surface marker expressed at native levels and without relying on a cell-type specific promoter we turned to L1CAM. L1CAM is an axonally-targeted protein that belong to a larger class of immunoglobulin-like cell adhesion molecules that mediate cell-to-cell recognition and are responsible to multiple aspect of neuronal development[134]. L1CAM is expressed on most neurons in the central nervous systems at the onset of differentiation, but is not expressed in glia cells. Using the same batch of tropism-null virus that we had used in the previous experiments, we formed anti-L1CAM-AAV composites. Composite virus carried tdTomato driven from the widely active CAG promoter. After two days *in vitro*, neuron cultures were infected, and after an additional 7 days fixed and stained for a neuronal marker (NeuN) and a glia marker (GFAP) (**Figure 2.7a**). Comparing immunohistochemistry from these experiments to controls (neuron culture infected with tropism-null HUH-AAV carrying CAG-tdTomato), we found that composite virus preferentially infected neurons by a large margin (**Figure 2.7b**). Whereas in control experiments neurons and glia were infected to similar degree (ratio ~1), with anti-L1CAM-AAV composites neuron infection was ~7.5-fold more likely. Overall, these data suggest that receptor-mediated viral gene delivery is possible with a mixture of on-target and off-target cells and without overexpression of the target receptor.

Discussion

We have demonstrated that AAV-DJ VR IV, which has been described in the context of fluorescent proteins insertion[51], is amenable to insertion of targeting scaffolds (including nanobodies and phage protein scaffolds) and that these scaffolds redirect tropism toward their cognate receptors. Unlike VR VIII, which is often the target of viral display studies[46], VR IV can tolerate large insertions (up to 30kDa). Unlike modifications of the VP2 N-terminus, another previously described way to engineer AAV tropism[41], there is little chance for the inserted targeting moiety to become trapped in the capsid interior. We showed that insertions can be made in capsid proteins all three capsid proteins independently, although modifications of VP1 & VP2 generally were better tolerated. An additional upside of selectively modifying VP1 is that if the modified capsid protein is not incorporated, the resulting viral particles will be non-infective. This is because VP1 contains a phospholipase domain that is critical for efficient endosomal escape[102]. When we consider that VP3 is ~10-fold more abundant in the AAV capsids than VP1, it is tempting to raises the possibility of a single viral capsid carrying multiple targeting moieties with different copy numbers. However, we have observed that AAV with modifications of VP3 (nanobody, Gp2, or mMobA HUH) can only be produced at high titer when unmodified VP3 is expressed in *trans*. This supports the idea that having too many copies of targeting scaffolds per capsid impacts capsid assembly and/or stability which would result in a lower titer overall, or could increase the fraction of empty particles that are produced (which would also appear

to lower the qPCR-determined titer). In addition, capsid engineering might introduce greater heterogeneity of recombinantly produced AAV particles. For example, not every AAV particle may incorporate a copy of an engineered capsid protein, which would manifest as apparently normal production titers but lower infectivity. Our Western blot and pulldown data do suggest that this is the case in our preparations, in particular for HUH-AAV. Further studies, ideally in several AAVs serotype, that quantitively assess the relationship between different capsid modifications, virus yield, stability, and function are required to better understand the tradeoff that are being made when AAVs are engineered in this fashion. This may help us to tailor targeting scaffold insertion strategies to maximize specificity of the engineered tropism while minimally impacting other viral properties.

Cell surface glycans and other co-receptors (often receptor tyrosine kinases) have been identified as the primary receptor cell attachment for AAVs[96,100], while the AAV receptor (AAVR) is critical for endosomal entry[35] that can occur through clathrin-dependent, clathrin-independent pathways, caveolar endocytosis, and others[135]. While usage of AAVR is conserved for most AAV serotype in primates, AAV can transduces at low level in the absence of AAVR, both *in vitro* and *in vivo*[136]. Together, this suggests that multiple uptake pathways exist for AAVs. An open question is whether viral uptake pathway preference is altered as a result of the introduced targeting scaffolds. It is conceivable that the introduction of targeting scaffolds simply switches the primary attachment step from glycans to the targeted receptor and that this interaction follows a simple

equilibrium model with targeting scaffold/receptor affinity determining probability of virus uptake. This would explain why tropism-retargeting maps directly to the receptor molecule recognized by the targeting scaffold added to the AAV capsid (e.g., GFP with anti-GFP-AAV, etc.) and why transduction efficiency is sensitive to surface receptor expression levels (e.g., hIR with $Gp2^{IR}$-AAV). It is also possible that uptake mediated by this alternate receptor directs modified AAV variants into different endosomal compartments that have better microtubule-dependent trafficking to the nucleus or from which AAVs has an easier time to escape. This would explain how AAV composites can have higher transduction efficiency in cell lines that less permissive to wildtype AAV infection (e.g., Jurkats infected with AAV-DJ vs. anti-CD7-AAV). The apparent resistance to neutralizing antibodies that is provided by introduction of nanobody could also be explained by this mechanism. In addition to interfering with cell attachment, neutralizing antibodies against AAV can interfere with the cellular trafficking mechanism that lead to perinuclear accumulation nuclear uptake[137]. Utilizing different receptors, whose identity influences sorting of the endocytosed cargo[138,139], may be a way to override the mistrafficking effects of neutralizing antibodies.

Most importantly, we demonstrate that HUH tag adapter domains mediate covalent attachment of ssDNA-modified antibodies to AAV. By reusing the same batch of HUH-AAV to form AAV composites that infect different cell types, we further advance the previously proposed concept[77,119] that adapter-containing AAVs can be produced once, bio-banked, and ‘armed’ with different targeting

scaffolds as needed. This represents a significant improvement to other approaches that incorporate targeting scaffolds in the viral capsid through genetic engineering. They still require viral vector packaging for each new target, so the cost-savings that are achievable when the base capsid is re-used can be significant. Furthermore, re-usable tropism-null AAV technology paves the way for a generalized re-targeting method toward arbitrary cellular receptors. One way this might be achieved is to leverage the large diversity of existing antibodies. Extensive antibody discovery pipelines have been established in the last decade that supported the development of therapeutic antibodies and antibody-drug conjugates[140]. It is conceivable that being part of antibody-AAV composites represents another mode of action for these already developed and validated targeting scaffolds. In this respect, the ability to prepare AAV composites with targeting scaffolds that don't have to be genetically engineered (e.g., adding a split-intein) is tangible advantage of the approach presented here. The modification of antibodies with ssDNA is an established step in immuno-PCR and uses off-the-shelf reagents and processes. Lastly, copper-free click chemistry can be applied to non-proteinaceous molecules as well. This means that other types of bio-polymers (DNA, polyethers, polysaccharides, etc.) are in theory amenable to programmable HUH-mediated assembly with AAV.

Most importantly, structurally separating the two functions of AAVs (targeting and payload encapsulation) means that both functions can be engineered independently. For example, one could alter the immunogenic profile

of the tropism-null base capsid to increase efficacy of gene delivery in the presence of neutralizing antibodies[7] without affecting the fidelity of the targeting mechanism. Improved versions of HUH tags can replace the ones we currently use, resulting in the next generation of HUH-AAV that exhibit higher coupling efficiency of the conjugated antibodies, improved capsid incorporation of the HUH-tag, a lower off-target infection. Other methods of viral retargeting (specifically, viral display[46], serotype reshuffling[39], and related directed evolution approaches) are inherently limited in this respect because structural components that mediate both functions (i.e. the capsid proteins) are shared, which might result in adaptive conflict when the same structure/sequence cannot simultaneously satisfy different constraints or results in loss of a function that was not under selection pressure.

Competing Financial Interests

The authors declare no competing financial interests.

Funding Sources

This study was supported by an NIH NIMH U01 MH109038 grant to D.S. and an NIH NIGMS R35 GM119483 grant to W.R.G. A.Z. and E.J.A. received salary support from a Biotechnology Training Grant NIH T32GM008347.

Acknowledgements

We thank Klaus Lovendahl for technical assistance and advice. Viral vectors used in this study were prepared by the University of Minnesota Viral Vector and

Cloning Core. We thank Kelly LaPara and Douglas Yee for providing the HEK293-FT cells stably expressing hIR/GFP or GFP alone, and Benjamin Hackel for providing the sequence of anti-hIR Gp2, as well insightful discussion.

Figure 2.1

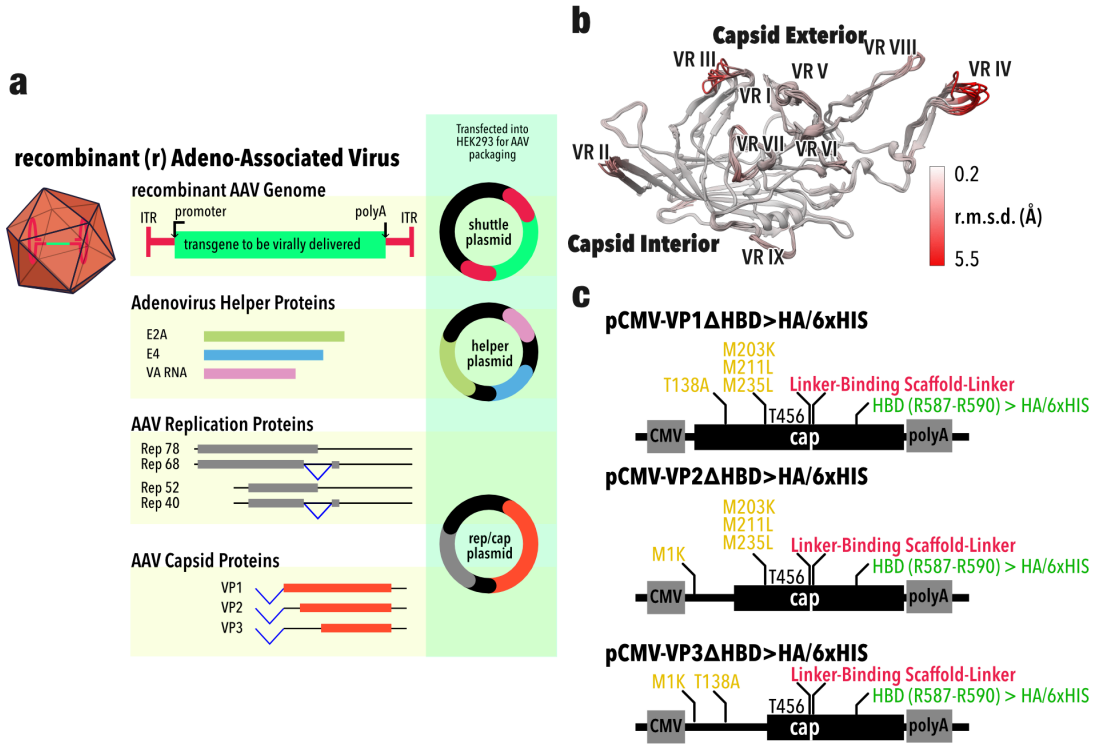


Figure 1. Engineering of AAV capsid protein. a, Recombinant AAV is packaged in HEK293 cells after transfection of three plasmids: the shuttle plasmid (contains the AAV payload, green) flanked by AAV2 ITRs, red); a helper plasmid encoding Adenovirus proteins (E2A, E4) plus VA RNA required for AAV replication and packaging; a plasmid that encodes AAV *rep* and *cap* genes. While the former expresses replication factors (e.g. Rep78), the latter gives rise to three capsid proteins (VP1-3) through alternative splicing (blue lines) and start codons (ORFs shown in orange). **b**, Superimposed crystal structures of VP from different AAV serotypes (AAV1 PDB 3NG9, 5EGC; AAV2 PDB 1LP3; AAV3 PDB 3KIC; AAV6 PDB 3SHM, 4V86, 3OAH; AAV8 2QA0; AAV9 3UX1). Root mean square deviation (rmsd) is mapped onto the structures (white to red). Variable regions (VR) and loops are annotated. **c**, Expression of individual modified capsid proteins is achieved by mutating (alternative) start codons (gold), inserting the targeting scaffold into position T456 (red), and replacing the endogenous heparan binding domain (HBD) with an HA or 6xHIS tag (green).

Figure 2.2

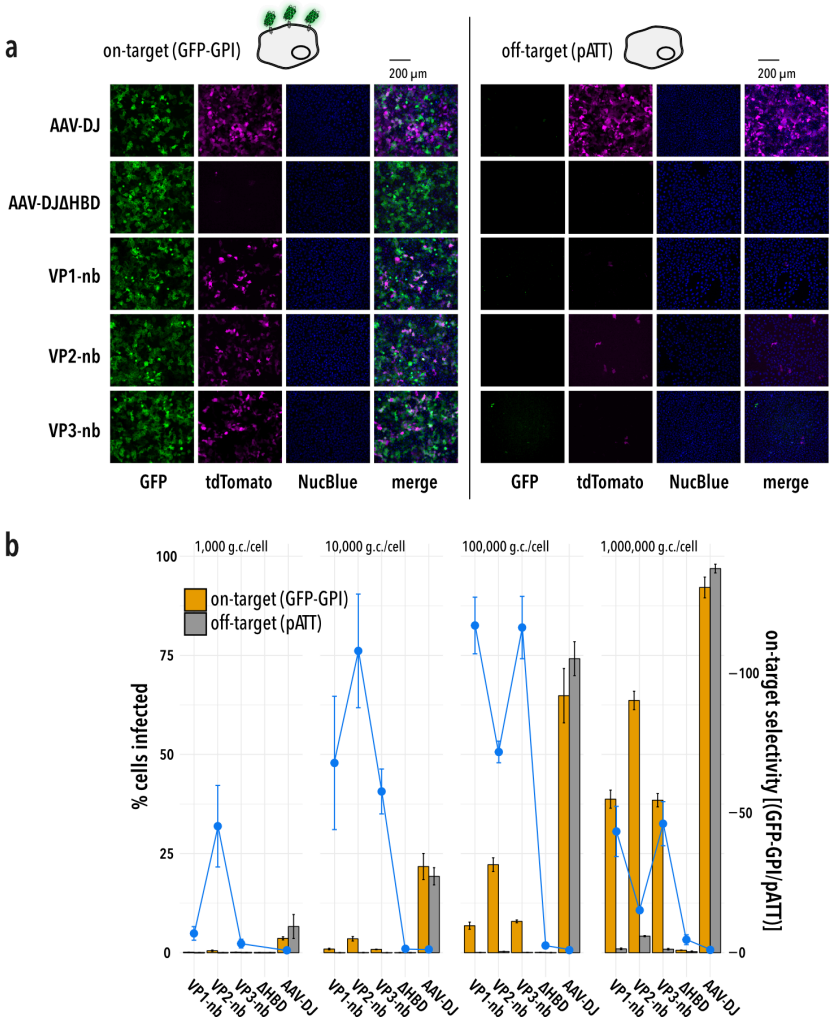


Figure 2. Receptor-mediated infection in a synthetic system. a, Wide-field fluorescent imaging of cells transiently transfected with surface-anchored GFP (left panels) or mock-transfected (right panels) 48 hours after viral transduction with $1x10^6$ g.c./cell. Individual channels are pseudo-colored green (GFP), magenta (tdTomato, the viral payload), and blue (Hoechst 33342). While wt AAV-DJ infects non-specifically, and AAV-DJΔHBD>HA infects neither, AAV-nb composites, in which the anti-GFP nanobody is incorporated into either VP1, VP2, or VP3, infects only GFP positive cells. **b**, On-target infection (GFP-GPI transfected, yellow) and off-target infection (mock transfected with plasmid pATT, grey) as a function of multiplicity of infection (MOI). On-target selectivity expressed as the ratio of on-target vs. off-target infection (blue dots, no selectivity = 1). All error bars indicate standard error (n = 3).

Figure 2.3

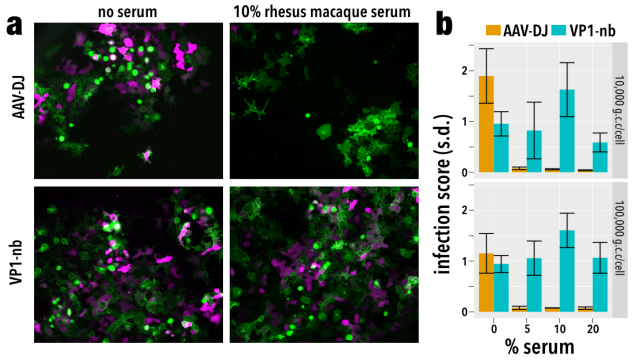


Figure 3. Nanobody-AAV are resistant to neutralizing antibodies. a, HEK293 expressing GFP (green) as a synthetic receptor on the extracellular face of the cell membrane were infected – in the absence or presence of 10% pooled, mixed gender rhesus macaque serum (BioIVT) – with either 100,000 genome copies (g.c.) / cell AAV-DJ or VP1-nb(anti-GFP) AAV composites, both delivering tdTomato (magenta). **b**, Infection scores for AAV-DJ (gold) and VP1-nb-AAV (teal) at different serum concentrations were determined by flow cytometry at two different multiplicities of infection.

Figure 2.4

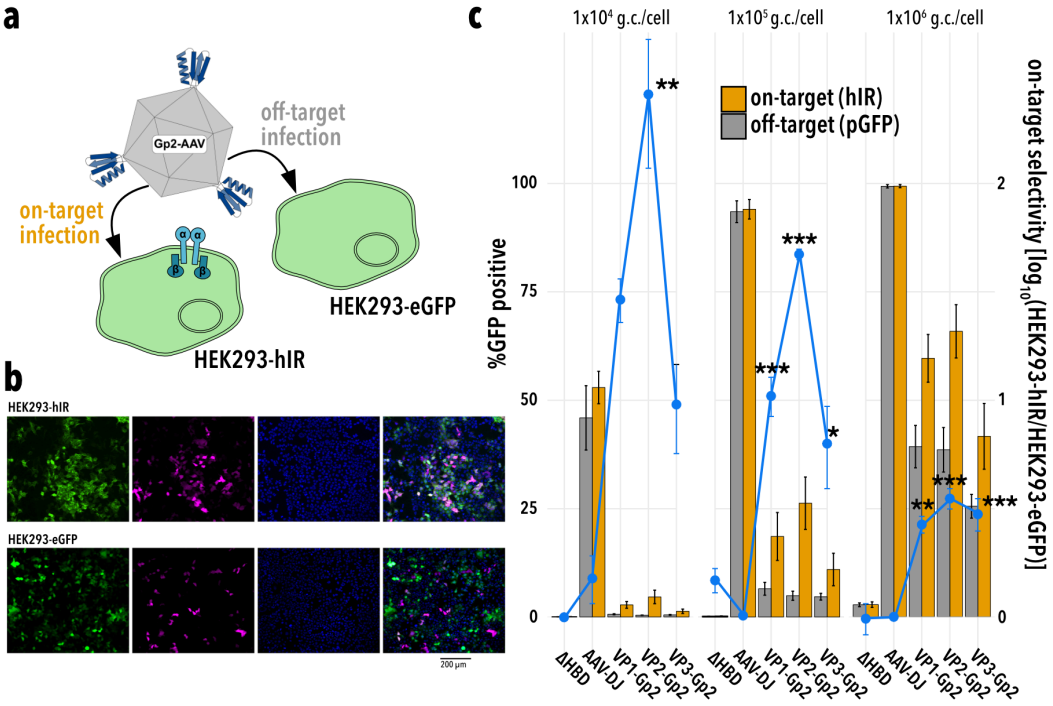


Figure 4. Receptor-mediated transduction of a breast cancer model cell line. a, AAV engineered to express a human insulin receptor (hIR) specific Gp2 targeting scaffold (Gp2 ^{IR}, blue cartoon representation) is expected to infect hIR-expressing HEK293 cells (on-target; also express GFP) and not control HEK293-EGFP cells (off-target). **b**, Wide-field fluorescent imaging 48 hours after viral transduction with 1x10^6 genome copies (g.c.)/cell VP2-Gp2^{IR}-AAV. Individual channels are pseudo-colored green (GFP), magenta (tdTomato, the viral payload), and blue (Hoechst 33342). Gp2 ^{IR} -AAV preferentially infects hIR-positive cells. **c**, Comparing AAV tropism for HEK293-hIR (orange), and HEK293-EGFP (grey). Bar height represents %infected cells at a specific multiplicity of infection (MOI, g.c./cell). While AAV-DJ is infecting cells indiscriminately, and AAV-ΔHBD in non-infective, Gp2 ^{IR} -AAV preferentially infects hIR over-expressing cells. The observed off-target infection likely is due to basal expression of endogenous hIR in control cells. Importantly, while off-target transduction is observed at high MOI, at lower MOI better specificity can be achieved. For example, at a MOI of 1x10^4 g.c./cell, the ratio of on-target vs. off-target infection is >11-fold for VP2-Gp2 ^{IR} -AAV. Significance of difference is on-target vs off-target infection is tested by a two-sided Dunnett's-test for multiple comparison with AAV-DJ as the control. Significance levels: * p<0.05, ** p<0.01, *** p<0.001, n.s. otherwise. All error bars indicate standard error (n = 3).

Figure 2.5

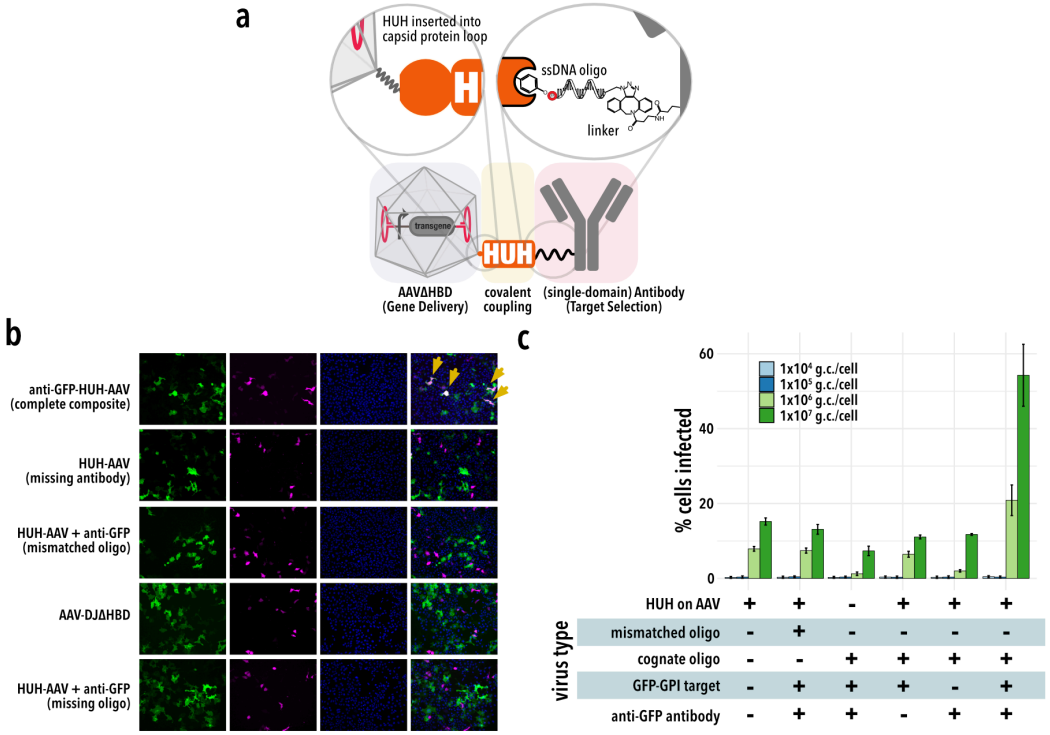


Figure 5. Generalized receptor-mediated infection using antibody-AAV composites. a, Covalently linked composite structures contain an AAV particle (packaging the delivered transgene) and an antibody that selectively binds cell-type specific surface markers. A HUH domain (here, mMobA) bridges the two component and confers programmable assembly. Insets provide further detail on HUH domain display on the capsid surface and HUH/antibody linkage (formation of a phosphotyrosine adduct with conjugated ssDNA-mAb). **b**, Wide-field fluorescent imaging of cells transiently transfected with surface-anchored GFP 48 hours after viral transduction with $1x10^6$ g.c./cell of the indicated virus. Individual channels are pseudo-colored green (GFP), magenta (tdTomato, the viral payload), and blue (Hoechst 33342). Only complete anti-GFP-HUH-AAV composite mediate specific infection of GFP positive cells (yellow arrows). With any component missing (mMobA HUH, oligo, or antibody) only basal off-target infection is observed, which is not specific to GFP positive cells (i.e., there is no colocalization of tdTomato and GFP signals). **c**, Infection efficiency is quantified at different multiplicities of infection (MOI) for the indicated virus. Bar height represents %infected cells at a specific multiplicity of infection (MOI, g.c./cell). All error bars indicate standard error (n = 3).

Figure 2.6

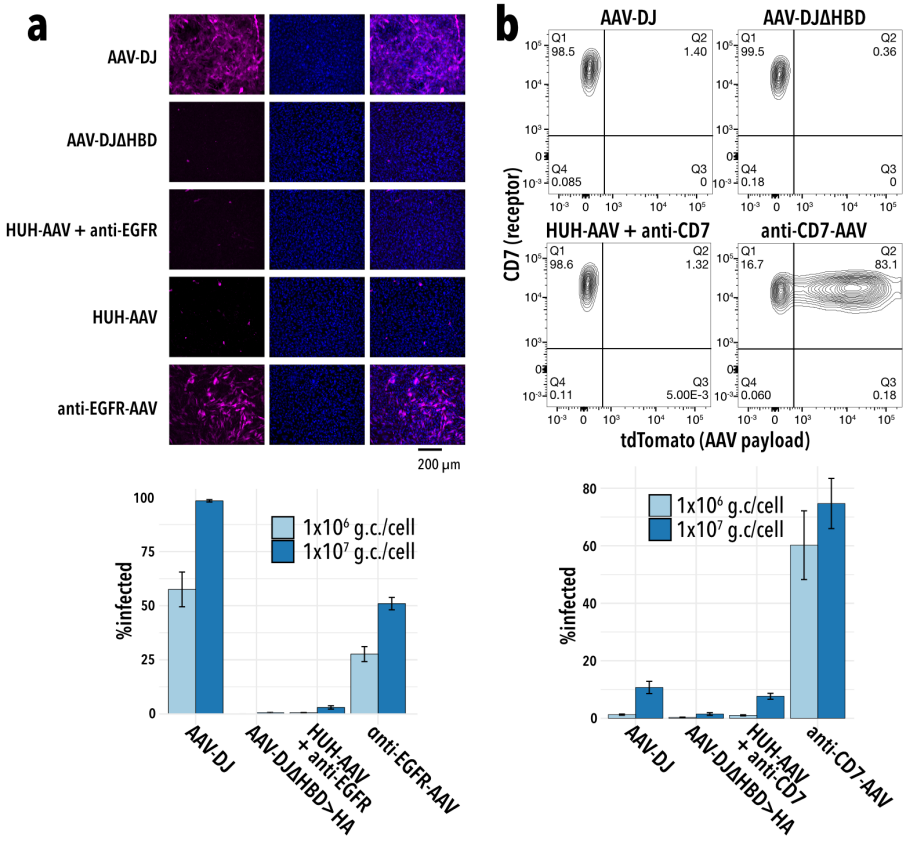


Figure 6. Efficient re-targeting of pre-made AAV. a, Wide-field fluorescent imaging 48 hours after viral transduction of U251 MG cells with 1x10^6 genome copies (g.c.) / cell of the indicated virus. Individual channels are pseudo-colored magenta (tdTomato, the viral payload) and blue (Hoechst 33342). While AAV-DJ and complete anti-EGFR-AAV composites infects strongly, AAV-DJΔHBD>HA and composites formed omitting either oligo or antibody are non-infective. Infection efficiency is quantified at different multiplicities of infection (MOI) for the indicated virus (bottom panel). **b**, Anti-CD7-AAV composites infect Jurkat cells. Shown are representative flow analysis examples from Jurkat cells infected with the AAV-DJ, AAV-DJΔHBD>HA virus, HUH-AAV composites formed with non-oligo labelled anti-CD7, and complete anti-CD7-AAV composites. Cells were infected with each virus at a MOI of 1x10^7 g.c. / cell. Cell are counterstained for surface CD7 expression. Percentage of cells in each quadrant are indicated in each corner. Only complete anti-CD7 -AAV composites transduce Jurkat cells efficiently. Infection efficiency is further quantified at different MOI (bottom panel). All error bars indicate standard error (n = 3).

Figure 2.7

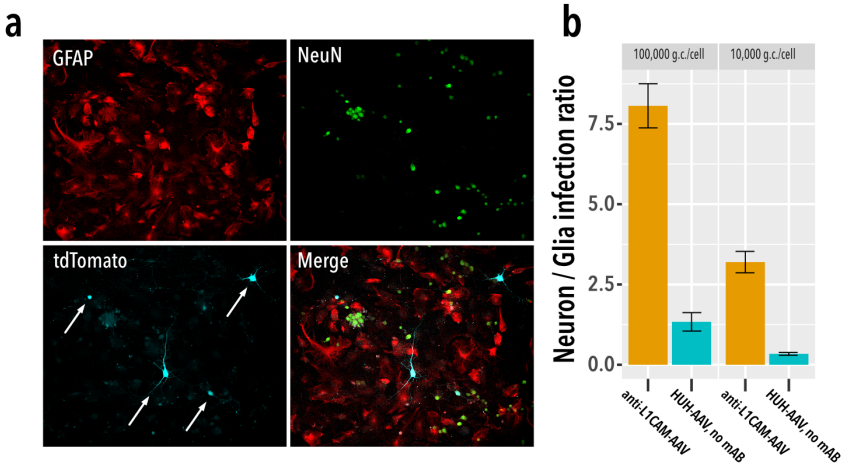
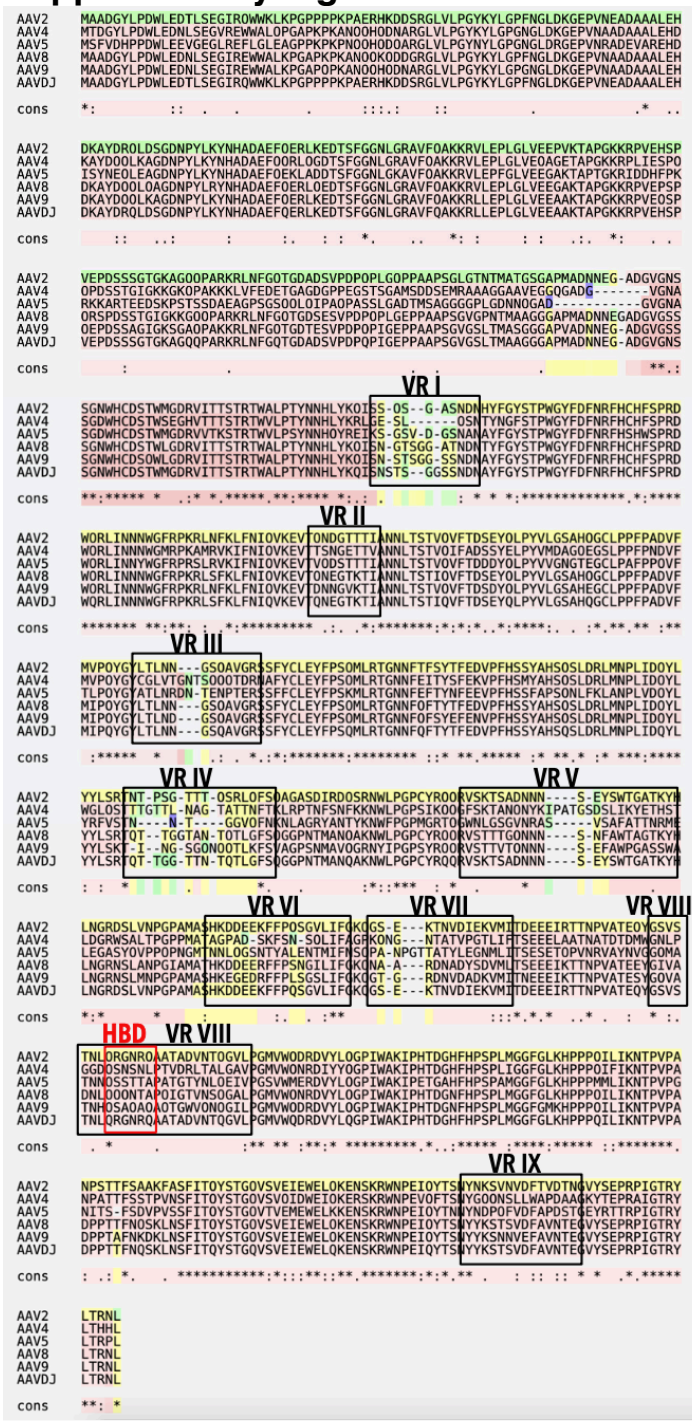


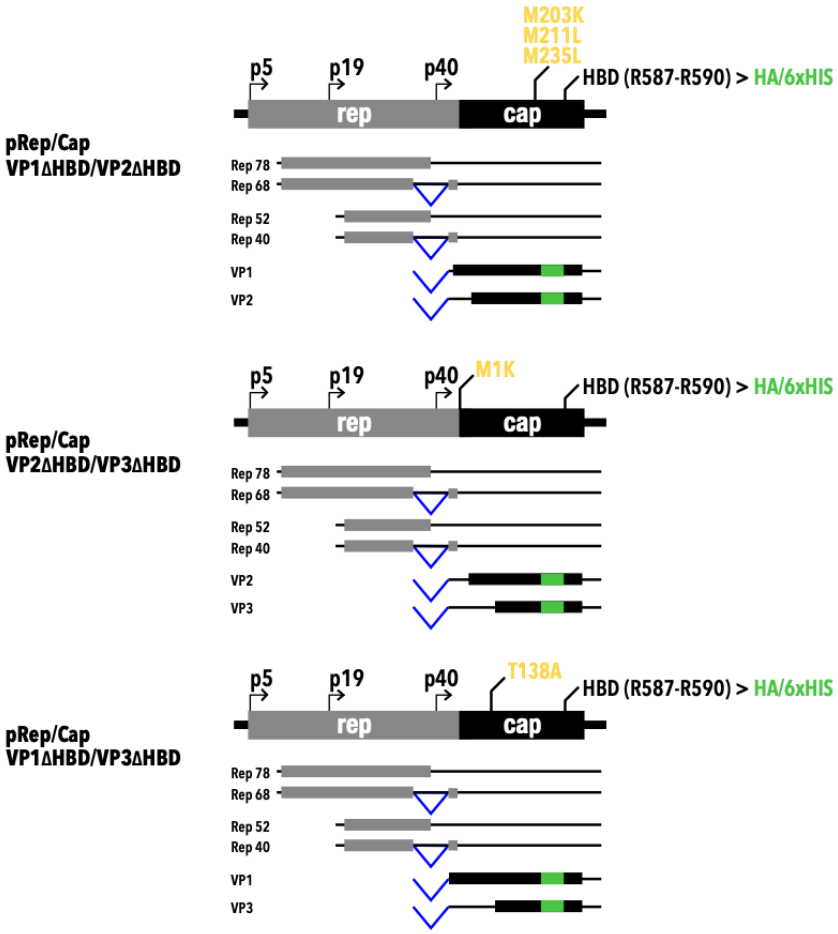
Figure 7. Receptor-mediated infection in primary neuron culture. a, Immunohistochemistry of primary neuron/glia co-culture infected with L1CAM-conjugated AAV delivering tdTomato. IHC identifies neurons (NeuN, green), glia (GFAP, red) and infected neuron (cyan, white arrows). **b**, Infection ratio of neuron vs. glia. While HUH-AAV without conjugated L1CAM antibody (HUH-AAV, no mAB) infects both neurons and glia (teal), L1CAM-conjugated AAV predominantly infects neurons (yellow). All error bars indicate standard error (n=7-9). g.c., genome copies; mAB, monoclonal antibody.

Supplementary Figure 2.1



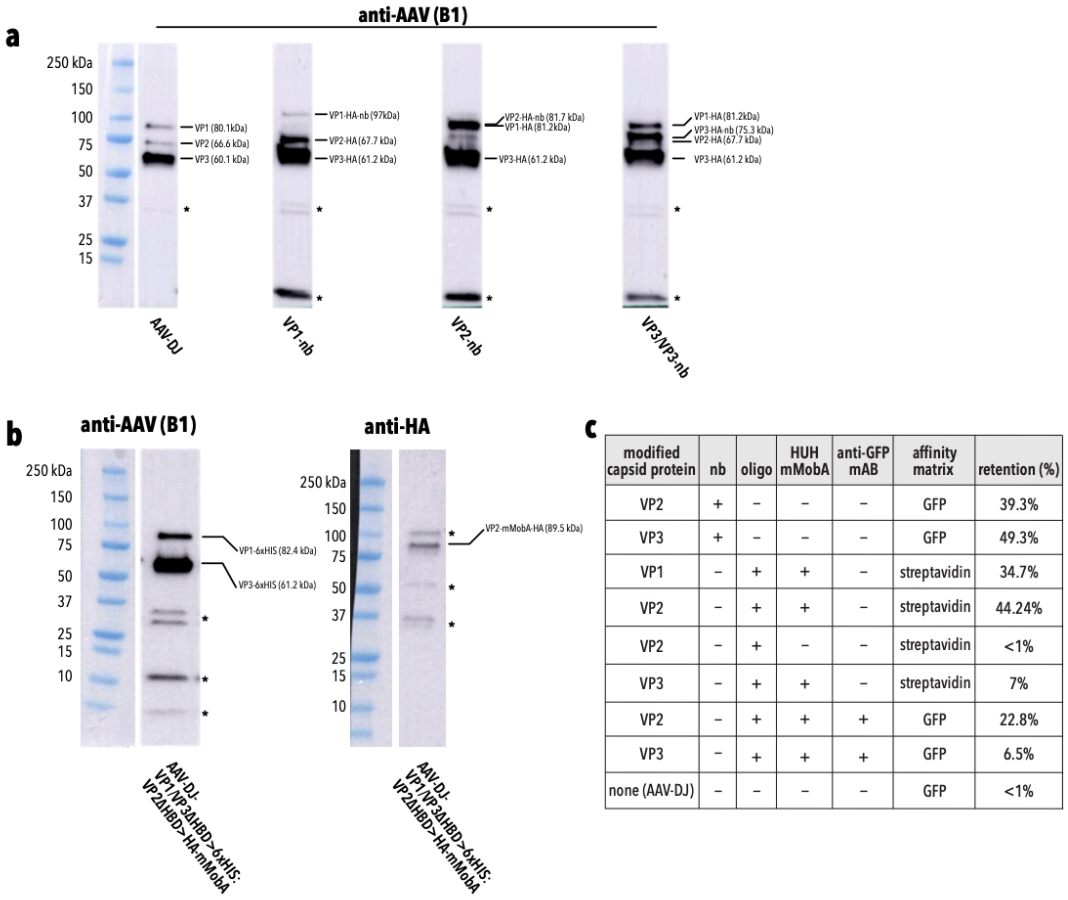
Alignment of AAV capsid proteins from different serotypes. Variable Regions (VR), the heparin binding domain (HBD), common loops are indicated by boxes.

Supplementary Figure 2.2



Trans-complementation plasmids for production of AAV with modified capsid proteins. Line drawing represent pAAV-DJ *rep* and *cap* genes, with endogenous promoters (arrows), and introduced mutations (start codons, yellow; heparin binding domain replacement with a HA tag, green). The resulting gene products expressed from each rep/cap variant are shown below.

Supplementary Figure 2.3



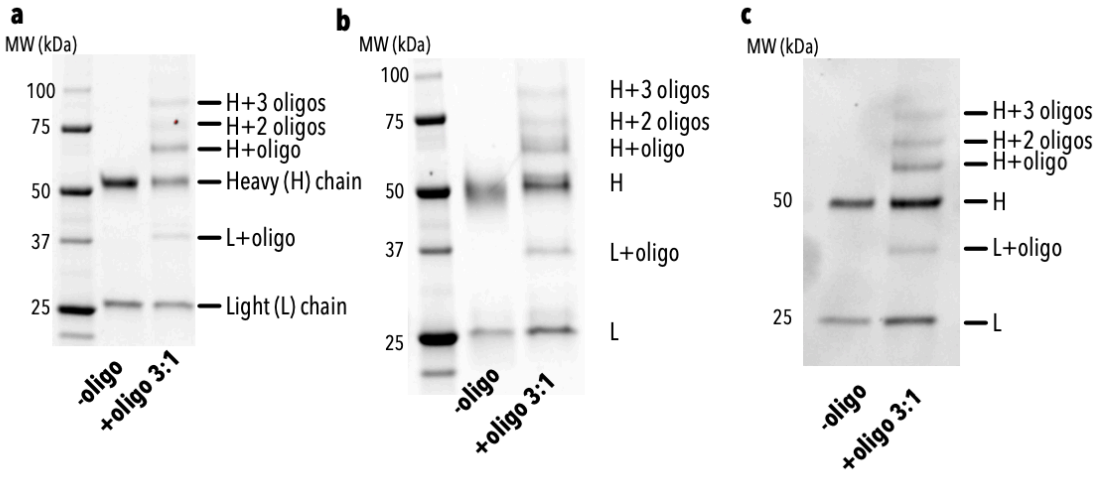
c

modified capsid protein	nb	oligo	HUH mMobA	anti-GFP mAB	affinity matrix	retention (%)
VP2	+	–	–	–	GFP	39.3%
VP3	+	–	–	–	GFP	49.3%
VP1	–	+	+	–	streptavidin	34.7%
VP2	–	+	+	–	streptavidin	44.24%
VP2	–	+	–	–	streptavidin	<1%
VP3	–	+	+	–	streptavidin	7%
VP2	–	+	+	+	GFP	22.8%
VP3	–	+	+	+	GFP	6.5%
none (AAV-DJ)	–	–	–	–	GFP	<1%

Western blot of nanobody- and HUH-modified capsids, and affinity pulldown of nanobody-or mMobA-modified virus. a, Gradient-purified AAV capsid proteins, either wildtype AAV-DJ, or AAV-DJΔHBD>HA containing anti- GFP-nanobody-modified VP1-, VP2-, or VP3 were separated by SDS-PAGE following Western Blot with the an anti-AAV antibody (clone B1). Capsid protein corresponding with each band, along with the predicted molecular weight, are shown. While the shift in molecular weight is noticeable for VP1-nb, VP2-nb and VP3-nb overlap with VP1 and VP2, respectively. Note that addition of the HA tag appears causes proteins to run slightly higher than their theoretical size and that

to produce VP3-nb, unmodified VP3 had to be expressed as well. * denotes non-specific bands. **b**, Gradient-purified VP2-mMobA-AAV capsid proteins (composition: AAV-DJ-VP1/VP3ΔHBD>6xHIS and VP2ΔHBD>HA-mMobA) were separated by SDS-PAGE following Western Blot with the an anti-AAV antibody (clone B1) or anti-HA antibody (clone 16B12). Capsid protein corresponding with each band, along with the predicted molecular weight, are shown. * denotes non-specific bands. **c**, Incorporation of nanobody or mMobA modified capsid protein was further investigated by affinity pulldown experiments with GFP-conjugated magnetic beads (to which anti-GFP-nb-AAV and anti-GFP-HUH-AAV bind) or Streptavidin-conjugated beads (bindingmMobA-AAV reacted with a biotionlylated mMobA-specific ssDNA oligo). Beads were washed, viral genomes were extracted from bead-bound capsids and analyzed using qPCR. Percent of affinity matrix-retained virus, compared to input material, is shown.

Supplementary Figure 2.4



Antibody conjugation with ssDNA oligo. anti-GFP IgG**(a)**,anti-EGFR IgG **(b)**, and anti-CD7 **(c)**conjugated to an HUH- specific oligo using a commercial copper-free click chemistry kit. Conjugation was carried out per kit instructions to achieve a ratio of 3:1 oligo:antibody. Unreacted and conjugated antibody were analyzed using denaturing SDS-PAGE. Average ratio of oligo:antibody was calculated by quantifying intensity of each band using FIJI image software. Experimental oligo:antibody ratio was calculated to be 2.29 for anti-GFP, 1.83 for anti-EGFR, and 1.40 for anti-CD7.

Supplemental Table 2.1

Virus Description	Capsid Constructs Used	Typical Titer GC/uL
VP1-nanobody	BP-00238, BP-00395	$6.79x10^{10}$
VP2-nanobody	BP-00142, BP-00396	$5.05x10^{10}$
VP3-nanobody	BP-00382, BP-00391, BP-00397	$4.54x10^{10}$
VP1-Gp2	BP-00238, BP-00447	$9.64x10^{11}$
VP2-Gp2	BP-00142, BP-00448	$7.88x10^{11}$
VP3-Gp2	BP-00382, BP-00391, BP-00449	$7.63x10^{11}$
VP1-mMobA	BP-00238, BP-00398	$1.37x10^{11}$
VP2-mMobA	BP-00142, BP-00437	$8.40x10^{11}$
VP3-mMoba	BP-00382, BP-00391, BP-00438	$8.51x10^{11}$
AAV-DJ	pAAV-DJ	$3.57x10^{11}$
AAV-DJ∆HBD>HA/6xHis	BP-00385, BP-00142	$8.22x10^{11}$

Typical titers for recombinantly produced AAV

Supplemental Table 2.2

Antigen	Clone	Source	Part Number	Application
GFP	9F9.F9	Abcam	ab1218	Attached to virus
EGFR	EGFR1	Abcam	ab30	Attached to virus
CD7	eGio124-1D1	eBiosciences, ThermoFisher	48-0079-42	Attached to virus
L1CAM	555	R&D Systems	MAB5674	Attached to virus
CD7-APC	eBio124-1D1, APC	eBiosciences, ThermoFisher	17-0079-42	Flow cytometry
EGFR	EGFR-PE	ThermoFisher	MA5-28544	Flow cytometry (not shown)
AAV	B1	ARP	03-61058	Western blot
HA-11	16B12	Biolegend/Covance	MMS-101P	Western blot
Goat anti-mouse secondary antibody HRP	Polyclonal	ThermoFisher	31430	Immuno-histochemistry
NeuN	EPR12763	Abcam	ab177487	Immuno-histochemistry
GFAP	2.2B10	ThermoFisher	13-0300	Immuno-histochemistry
Goat anti-rat IgG cross-adsorped secondary antibody, Alexa Fluor 350	Polyclonal	ThermoFisher	A-21093	Immuno-histochemistry
Goat anti-Rabbit IgG cross-adsorbed secondary antibody, Alexa Fluor 700	Polyclonal	ThermoFisher	A-21038	Immuno-histochemistry

Commercial antibodies used in this study

Chapter 3:

Split *Staphylococcus aureus* prime editor for AAV delivery

Eric J. Aird[1,2], Alina C. Zdechlik[1,2], Brian L. Ruis[1,2], Colette B. Rogers[1,2], Andrew L. Lemmex[1,2], Andrew T. Nelson[1,2], Eric A. Hendrickson[1,2], Daniel Schmidt[2,3], and Wendy R. Gordon[1,2*]

[1]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, MN, USA

[2]Center for Genome Engineering, University of Minnesota, Minneapolis, MN, USA

[3]Department of Genetics, Cell Biology, and Development, Minneapolis, MN, USA

DOI: https://doi.org/10.1101/2021.01.11.426237

Author Contributions

E.J.A. and A.C.Z. conceived idea and designed experiments. E.J.A., A.C.Z., B.L.R., C.B.R., A.L.L, and A.T.N. carried out experiments and analyzed data. A.C.Z. assisted with molecular cloning, cell cytometry, and creation of the L202S cell line. W.R.G, D.S., A.C.Z., and E.A.H. advised and aided in assembly of the manuscript. E.J.A. authored the manuscript with input from all authors.

Preface

Prime editing brings immense promise to correct a large number of human pathogenic mutations and enact diverse edit types without introducing widespread undesired editing events. Delivery of prime editors in vivo would enable such edits to be introduced in a clinical setting. The coding sequence for prime editor, however, is too large to fit within the size-constrained adeno-associated virus (AAV) genome. Herein, we describe a split *Staphylococcus aureus* prime editor capable of being delivered by dual AAVs. We characterize the editing ability of plasmid-based versions of an *S. aureus* prime editor in vitro at a variety of loci with diverse edit types. We investigate various split prime editor architectures and alternative dimerization domains. Finally, we demonstrate the capacity of prime editor to be co-delivered by dual AAVs in vitro. While editing rates are lower than desired, this approach presents an important step to translate prime editing for in vivo delivery.

Introduction

Genome editing has brought incredible promise to correct or ameliorate previously untreatable genetically linked diseases such as Tay-Sachs disease or Phenylketonuria (PKU). The use of programmable nucleases such as CRISPR-Cas9 that introduce a DNA double-stranded break (DSB) enables precise gene editing at a desired locus[141,142]. Traditionally, homology-directed repair (HDR) has been the only means to introduce precise nucleotide changes at a DSB via the supplementation of a donor DNA molecule encoding the changes. HDR, however, usually occurs less frequently than the circumstantially undesired non-homologous end joining (NHEJ) pathway[143]. To circumvent the NHEJ pathway, base editors were designed to change single bases in DNA without the need to create a DSB[144]. Base editors employ nickase Cas9s that have one of the nuclease domains mutated to prevent cleavage on one DNA strand. Base editors, as the name suggests, are limited to introducing transition point mutations (purine:purine or pyrimidine:pyrimidine), although recently C to G base editors have been characterized[145,146]. While NHEJ levels are low with base editors and high rates of precise editing can be achieved, the restraint on editing type and inability to edit specific bases within a window makes it not ideal for many desired precise genome editing applications.

The latest advent in CRISPR genome editing technologies is prime editor, which is capable of introducing a wide change of precise modifications without the

need for DSB formation[147]. Like base editor, prime editor utilizes a Cas9 (H840A) nickase. A reverse transcriptase (RT) is tethered to the nickase while a prime editing guide RNA (pegRNA) contains a 3' extension serving as both the genome hybridizing site (primer binding site; PBS) and the RT template encoding the desired edits. Prime editor cleaves the single strand of DNA, the PBS of the pegRNA hybridizes with the newly exposed ssDNA, and the RT synthesizes from the RT template. Flap resolution and DNA mismatch repair then allow for incorporation of the desired modification. Therefore, prime editing is theoretically able to make edits downstream of the nick site, termed the +1 site. Prime editor was shown to allow for edits ranging from codon changes to small insertions and deletions through encoding these modifications in the 3' pegRNA extension[147]. Hypothetically, prime editor is capable of correcting close to 90% of human pathogenic mutations[147]. As with other types of genome editing technologies, however, in vivo targeting and delivery remains a large hurdle to overcome to achieve such a goal[2].

Adeno-associated virus (AAV) is the most common in vivo delivery vehicle for gene editing reagents. AAV contains a single-stranded DNA (ssDNA) genome that can be replaced with transgenes of interest. The virus can target a wide range of tissue and cell types with a relatively low immunogenicity profile. AAV-mediated delivery of CRISPR-Cas has been successfully deployed to produce indels[148], create chimeric T-cell receptors[149], and generate large deletions[67]. Extensive engineering efforts have also modified the virus to enhance cell-specific delivery

or decrease immunogenicity[1]. A major constraint of AAV is the maximum size of its ssDNA genome, roughly 4.7 kb not including the two flanking inverted terminal repeats (ITRs)[31]. The conventional Cas9 originating from *S. pyogenes* (SpCas9) is 4.2 kb in size, allowing little space for necessary regulatory elements such as promoters and terminators and precludes encoding of a gRNA expression cassette in the same genome. To overcome this limitation, smaller orthologs of Cas9 such as from *S. aureus* (SaCas9) have been utilized[65]. The gene length of SaCas9 is approximately 1 kb shorter than SpCas9 allowing for more freedom in packaging in AAV. Another strategy to deliver full-length Cas proteins has been to split the protein into two AAV vectors to be co-delivered. This approach can take place on the DNA, RNA, or protein level (reviewed in [150]). On the protein level, engineered split trans-splicing inteins can be co-opted essentially as dimerization domains to bring two halves of a split protein together [151]. For base editor to be packaged and delivered by AAV, a split base editor, split intein system was used[64,66]. The two intein halves associate at low nM affinity and swiftly excise themselves out, leaving only a small scar. For both SpCas9 and SaCas9, various groups have identified numerous permissible split locations that allow the proteins to retain nearly full activity once recombined[64,152].

Prime editor was previously shown to be delivered ex vivo using a three-part lentivirus system to mouse primary cortical neurons[147]. This approach, however, is not broadly applicable in the clinic due to constraints on choice of cell type to edit. Ribonucleoprotein (RNP)-containing lentiviral particles have been used to

transiently deliver Cas9[153], but the sheer size of prime editor protein (~250 kDa) makes it difficult to produce conventionally in *E. coli* and package. Ideally, AAV would be used to deliver prime editor. At ~6.4 kb, prime editor is too large to be encoded into a single AAV genome. Even when constituted as a split version, prime editor is unable to fit into two AAV genomes with the necessary regulatory elements and the pegRNA expression cassette.

To address this size constraint issue, we developed a new version of prime editor using SaCas9(N580A) as the guided nickase. We demonstrate that *S. aureus* prime editor (SaPE) can also introduce targeted changes to a wide range of genomic loci from point mutations to larger deletions. We assessed three different split locations in concert with split intein and split NanoLuc approaches to reconstruct full-length, active protein. All versions of split SaPE were capable of inducing editing events, although they tended to be less efficient than the full-length protein. The split version can be packaged into two standard size AAV genomes and be delivered in viral form in vitro. This work opens a potential avenue for in vivo and clinical exploration of prime editing.

Materials and Methods

Nucleotide and amino acid sequences

The following tables contain sequences for pegRNAs (**Supplementary Table 3.1**), sequencing primers (**Supplementary Table 3.2**), and proteins (**Supplementary Table 3.3**). All oligonucleotides were synthesized by Integrated DNA Technologies (IDT).

Cloning and DNA assembly

All prime editor expression vectors were generated using NEBuilder HiFi DNA assembly (New England Biolabs; NEB). The pegRNA expression vectors were generated as described below from Addgene plasmid #132777 (Generously provided by David Liu). AAV expression vectors were generated using NEBuilder HiFi DNA assembly combined with standard restriction digestion of the AAV vector backbone pAAV-CAG-GFP containing standard AAV2 ITRs. The *S. pyogenes* prime editor expression vector was a gift from David Liu (Addgene #132775). *S. aureus* Cas9 was amplified from a vector courtesy of Feng Zhang (Addgene #61591). Q5 site directed mutagenesis (New England Biolabs) was used to generate the Npu (C1A) mutation. Assembly reactions were transformed into competent Stellar cells (Takara Bio). Plasmid DNA was purified either as minipreps (Qiagen) or maxipreps (Thermo Fisher). DNA concentration was quantified using

a Nanodrop (Thermo Fisher) and sequenced verified by Sanger sequencing (Genewiz).

pegRNA cloning

The pegRNAs were cloned using a protocol adapted from the Liu lab[147]. These modifications were made to incorporate the *S. aureus* gRNA scaffold sequence and to add some streamlined features in regard to vector digestion and golden gate assembly cycling conditions. A detailed protocol is provided in **Supplementary Note 3.1.** For *S. aureus* pegRNAs, the annealing protospacer oligonucleotides were designed as such: Forward (CACC ...(G) (spacer sequence)... GTTTT) and reverse (TACTAAAAC …(reverse complement) (C)...). Add a G prior to spacer sequence if it doesn't begin with a G (and corresponding C on the reverse oligonucleotide). The pegRNA 3' extension annealing oligonucleotides were designed as such: Forward (GAGA …(RT template + PBS)...) and reverse (AAAA …(reverse complement)...). The phosphorylated scaffold oligonucleotides were: Forward (/5Phos/ agtactctggaaacagaatctactaaaacaaggcaaaatgc cgtgtttatctcgtcaacttgttggc) and reverse (/5Phos/ tctcgccaacaagtt gacgagataaacacggcattttgccttgttttagtagattctgtttccagag). Golden gate assembly was performed with 1 μl each of 1 μM stocks of these 3 annealed oligonucleotides, 250 ng of pU6-pegRNA-RFP acceptor (Addgene #132777, courtesy of David Liu), 1 U of BsaI-HFv2 (NEB), 1 U of T4 DNA ligase (NEB), and 1x T4 DNA ligase buffer

in 10 μl total volume. Reaction conditions were as follows: 10 cycles of 5 min at 37 °C and 10 min at 16 °C followed by inactivation steps of 5 min at 55 °C and 5 min at 85 °C. 1 μl of the assembly reaction was transformed into competent Stellar cells. Non-red colonies were picked for subsequent DNA isolation.

Cell culture

HEK-293T, U2-OS, HCT116, RPE1, and HEK-293 cells were cultured in DMEM (Corning) supplemented with 10% FBS (Gibco) and 0.5% penicillin/streptomycin (Gibco). Cells were incubated at 37°C in 5% CO_2. Bxb1-mediated recombination was used to generate the stable, single copy GFP(L202S)-2A-mKate2 HEK-293T cell line[154].

Plasmid transfection

Cells were seeded in a 48-well plate at 30,000 cells per well or in a 24-well plate at 60,000 cells per well. Approximately 24 hr post-seeding, cells were transfected using Lipofectamine 2000 (Invitrogen). A 1:1 molar ratio of prime editor to pegRNA vector was used (250:83 ng in 48-well plates or 500:167 ng in 24-well plates) according to the suggested manufacturer's protocol. For split prime editor transfections, the total amount of prime editor vector was held constant. Cells were incubated for 72 hr post-transfection for all downstream analyses.

Analysis of reversion of GFP L202S mutation

Flow cytometry analysis was carried out on a BD Fortessa X-20 instrument at the University of Minnesota Flow Cytometry Resource. Cells were prepared by first washing the cells with PBS and detaching with Accutase (Sigma). Cells were gently pelleted, washed with ice-cold PBS, and resuspended in ice-cold PBS supplemented with 5% FBS. Data from 10,000 to 100,000 cells was collected using BD FACSDiva software and compiled using FlowJo (version 10.6). Cells were initially gated based on FSC-A and SSC-A (for live cells) and then gated on FSC-W versus FSC-H (for single cells). The presence or lack of GFP expression was then evaluated (**Supplementary Note 3.2**).

In separate experiments, live cell imaging was carried out on an Olympus IX83 inverted microscope equipped with an Andor iXon Ultra 888 EM-CCD. Fluorescence was provided by a Sola light engine (Lumencor). For bioluminescence imaging, a Semrock light filter FF01-460/60 was used to capture NanoLuc emission. Images were processed using Fiji (version 1.51r).

Next generation sequencing

Genomic DNA was isolated 72 hr post-transfection using the Quick-DNA Miniprep Plus kit (Zymo Research) and eluted into 25 µl 10 mM Tris-HCl, pH 8. A 150-250 bp region encompassing each targeted locus was PCR amplified from ~40 ng genomic DNA with ends containing partial Illumina adapter sequences using CloneAmp HiFi PCR (Takara). Reaction conditions were as follows: 98 °C for 1 min, then 30 cycles of 98 °C for 10 sec, 55 °C for 10 sec, and 72 °C for 5 sec.

1 µl of each unpurified amplicon was then carried to a second PCR reaction using KAPA HiFi Library Amp (Roche Sequencing). NEBNext Multiplex Oligos (New England Biolabs) were used to add single indexes to the amplicons. Reaction conditions were as follows: 10 cycles of 98 °C for 20 sec, 61 °C for 15 sec, and 72 °C for 15 sec. 2 µl of each common amplicon were pooled and gel purified from a 1.5% agarose gel (Nucleospin clean-up, Takara Bio), eluting in 35 µl 10 mM Tris-HCl, pH 8. Common amplicon libraries were quantified with qPCR using NEBNext Library Quant kit (New England Biolabs) and pooled to equal concentrations. Sequencing was performed with an Illumina MiSeq with 2 x 150 bp paired-end reads (Genewiz). Sequencing reads were demultiplexed and analyzed using CRISPResso2 in batch mode[155].

GFP to BFP editing

HEK-293 cells stably expressing GFP were plated at 200,000 cells per well 24 hr prior to transfection. 250 ng of prime editor, 100 ng of pegRNA, and 50 ng of mCherry plasmid was then transfected using Lipofectamine 3000 (Invitrogen). For PE3 experiments, 100 ng of gRNA was also added. Cells were then incubated for 72 hr and analyzed using flow cytometry, gating for mCherry and BFP positive.

PIGA editing

PIGA^{-} cell lines (HCT116 clone #2D2 or RPE1 clone #1A10) were co-transfected with 500 ng pegRNA plasmid and 1.5 μg DNA of full-length SaPE or

750 μg each of split SaPE plasmids. $1x10^6$ cells were electroporated using the Neon Transfection System (1530 V, 10 ms, 3 pulses, 10 μL tips). Transfected cells were transferred to prewarmed media in 10 cm plates and incubated for 72 hours prior to collection for downstream analysis.

Western blot

Lysates were collected 72 hr post-transfection from HEK-293T cells in 24-well plates using RIPA buffer containing protease inhibitors. One third volume of each lysate was electrophoresed on a 4-20% SDS-PAGE gel and transferred to a nitrocellulose blot. The blot was blocked in 5% milk in TBS-T then incubated overnight at 4 °C with 1:1000 dilutions of primary antibody. Primary antibodies used were mouse anti-FLAG M2 (F1804; Sigma) or the loading control mouse anti-β-tubulin (T8328; Thermo Fisher). Blots were washed and then incubated for 1 hr with 1:10000 goat anti-mouse IgG-HRP (62-6520; Invitrogen). Blots were imaged using chemiluminescent buffer (Perkin Elmer) on an Amersham 600 UV imager (GE Healthcare).

AAV production

All viral vectors used in this study were generated by the University of Minnesota Viral Vector and Cloning Core (Minneapolis, MN). Briefly, AAV293 cells at 60% confluence were transfected with 600 µg of DNA (viral shuttle vector encoding the payload, helper plasmid, rep/cap plasmids at 1:1:1 ratio) using

polyethylenimine. 24 hr after transfection, the media was changed, and cells were checked for fluorescent protein expression (when applicable) to confirm transfection. 72 hr after transfection, cells were detached and pelleted. Viral particles were released from producer cells by repeated freeze/thaw cycles in the presence of Benzonase (100 units). Crude lysates were cleared by centrifugation and further purified using sucrose gradients. Viral particles in the supernatant were titered using qPCR with ITR-specific primers. Kanamycin-specific primers were used to confirm the absence of plasmid DNA after Benzonase treatment.

AAV transduction

HEK-293T cells were plated 24 hr prior to transduction at 50,000 cells per well. Cells were washed with 1x DMEM (no FBS) and AAV diluted in DMEM was added gently on top. Experiments were performed at ~1 × 10^6 g.c.(genome copies)/cell per virus. 1 hr after virus addition, 1 ml of D10 was added on top in each well. 24 hr post-transduction, media was aspirated and 500 μl fresh D10 was added. Cells were incubated a further 48-72 hr prior to analysis.

Results

Development of an S. aureus *prime editor*

As SaCas9 and SpCas9 share similar domain architecture, we reasoned that a nickase SaCas9 could serve as a sufficient nicking nuclease to be combined with the prime editing methodology. We generated the analogous HNH domain mutation in SaCas9 (N580A) as in SpCas9 (H840A) and tethered a reverse transcriptase to create an *S. aureus* prime editor, SaPE (**Figure 3.1a**). The previously engineered version of Moloney murine leukemia virus reverse transcriptase (eMMLV-RT) was utilized[147]. We then sought to create a split version of SaPE capable of fitting into two AAV genomes. Numerous methods have been applied to efficiently recombine split portions of Cas9[63,156]. We utilized a split intein-mediated approach wherein the split Npu trans-splicing intein from *Nostoc punctiforme* was appended onto the N- and C-termini of split SaPE[157] (**Figure 3.1a**). We trialed three different split locations in SaCas9*,* all of which have been previously reported[64,152]. Two of the split locations, E739/S740 (version 1) and K534/C535 (version 3), showed reasonable intein splicing in cells (**Figure 3.1b**).

To more quickly test editing conditions, we created a stably integrated fluorescent reporter cell containing a point mutation in *GFP(L202S)* that ablates GFP fluorescence linked to mKate2 via a 2A self-cleaving peptide[158] (**Figure 3.1c**). The restoration of GFP fluorescence through a +6 G to A transition point mutation can readily be detected using microscopy or flow cytometry (**Figure 3.1d**). We first tested full-length and split versions of SaPE alongside a variant with the RT

tethered to the amino terminus of nSaCas9 (PESa). In comparison with SpPE, SaPE had an 8-fold reduction in editing efficiency (**Figure 3.1e**). While SaPE editing efficiencies were low at around 0.5%, all split versions had comparable levels to full-length SaPE. These low editing frequencies with SaPE, as elaborated on in the Discussion, are a trend that generally holds true at most loci. Decreases in editing efficiency might be attributed to, among other factors, shorter residency time of SaPE on the DNA as opposed to SpPE[159]. Nonetheless, detectable levels of prime editing driven by SaPE are observed.

Prime editing necessitates optimization of the pegRNA 3' extension design as both the RT template and primer binding site (PBS) lengths appear to be edit type and locus dependent[147,160,161]. At the *GFP(L202S)* locus, we interestingly observed little difference in editing frequencies when sampling altered 3' extension designs (**Figure 3.1f**). Variation did exist among the three split versions, and in these sets of experiments, split SaPE had noticeable decreased editing frequencies compared to full-length. Combined with the intein splicing and editing profiles, we carried forward predominantly with version 3 of sSaPE (K534/C535).

In another fluorescent assay targeting a different locus in *GFP*, we once again compared SpPE to SaPE. A two amino acid change in GFP (T65S-Y66H) converts the fluorescence to BFP[162]. When prime editing reagents were transfected in HEK-293 cells stably expressing GFP, we observed a similar frequency of editing between full-length and split SaPE (1.85% versus 2.45%) (**Figure 3.1g**). This was again lower than SpPE (3.6%) but to a lesser degree than *GFP(L202S)* where

SpPE editing was almost an order of magnitude higher. Conversion to BFP was also seen in HCT-116 cells to a similar degree as in HEK-293 cells (**Supplementary Figure 3.1**). Interestingly, PAM sequence constraints required this edit to be upstream from the nick site (-1 site) in addition to a +2 point mutation. The ability of prime editor to make modifications upstream of the predicted nick site, 3 bp away from PAM, lends itself to past research suggesting Cas9 also can cleave 4 bp upstream[163]. While this effect might be locus dependent, we observed appreciable editing and potentially a broadened capability of prime editor.

Optimization of split SaPE

To try increasing editing frequencies at the *GFP(L202S)* locus with SaPE, we next sought to optimize experimental parameters and platform design. We performed a titration of plasmid DNA concentration using lipofection while the molar ratio of PE:pegRNA was held constant. Due to the PE vectors being approximately three times the size of the pegRNA plasmids in bp, this resulted in a 3:1 ng of DNA concentration ratio. The maximal editing of ~0.6% was seen at surprisingly the lowest concentration of DNA. (**Supplementary Figure 3.2**). In the case of sSaPEv3, higher DNA concentrations resulted in roughly a 33% decrease in editing efficiency. Cytotoxicity associated with large plasmids combined with excess amounts of DNA could be a cause of this trend[164].

Next, two previously described linkers were installed in place of the built-in one between nSaCas9 and RT to try enhancing editing rates: XTEN[165], found in

base editors, and $(H4)_2$, a rigid alpha helical linker of the sequence $(A(EAAAK)_4A)_2$ [166]. While the XTEN linker performed as well as the original SaPE linker, the replacement with the $(H4)_2$ linker ablated prime editing (**Supplementary Figure 3.2**). This difference highlights the crucial, but often ignored, impact of spatial orientation of components in fusion proteins. Further exploration of linkers and different permissible fusion locations of RT on nCas9 could be carried out to try optimizing this key parameter.

Another approach we utilized to try boosting editing rates was to employ the 3rd generation prime editor system, termed PE3[147]. PE3 uses a second gRNA that nicks the non-prime edited strand to encourage DNA repair machinery to preferentially repair the nicked strand once the desired edit has been incorporated. The hypothesis is to push the equilibrium of flap resolution and DNA repair towards the desired outcome, a method successfully used in base editors[144]. Due to the limited number of SaCas9 PAM motifs in the vicinity of the targeted location in *GFP(L202S)*, we could only assess two PE3 gRNAs in combination with SaPE. Use of either gRNA, which nick at -70 or -51 bp from the +1 site, resulted in lower editing efficiencies for sSaPEv3 (**Supplementary Figure 3.2**). For sSaPEv1, the -51 gRNA caused a dramatic tripling in editing efficiency (0.20% to 0.67%) while the -70 gRNA had little effect. PE3 has been shown to be moderately successful at improving prime editing at other loci with SpPE[147,167]. We chose, however, not to generally pursue the PE3 approach at other loci due to the limited amount of neighboring PAM sequences available. Taken together, our efforts to optimize

SaPE did little to improve editing rates from the outset in the context of the *GFP(L202S)* locus.

Broad assessment of SaPE

We next wanted to assess SaPE and sSaPEv3 across a wide range of loci and editing types while also changing the lengths of the 3' extension components of pegRNA. Using next-generation sequencing, we found SaPE capable of making wide-ranging precise changes to target loci (**Figure 3.2**). In sampling different RT template and PBS lengths at the *EMX1* locus in a +6 G to T transversion point mutation, we found variable editing rates between 0.3% and 2.1% for full-length SaPE (**Figure 3.2a**). For split SaPE, less variability existed, but editing rates were limited to between 0.3% and 0.5%. Split SaPE tended to have a decreased editing frequency across all loci tested, a trend that is not uncommon with split Cas9[62,63] but is in contrast to what was observed with split base editor[64]. At both the *FANCF* and *DNMT3B* loci, editing rates peaked around 0.4% for a range of point mutations, insertions, and deletions (**Figure 3.2b and c**). Point mutations, however, tended to yield higher editing efficiencies than insertion or deletions. In varying the RT template length at *RUNX1*, the longer length (15 nt) yielded improved editing over the shorter RT template (10 nt), regardless of the location of the edit from the nick site (**Figure 3.2d**). In the case of the further +7 G to A mutation, a 10 nt RT template resulted in nearly undetectable levels of editing.

The highest prime editing frequencies were seen at the *HEK3* locus, where a +3 C to A transversion was formed in 9.2% of full-length SaPE transfected loci and nearly 5% in the sSaPEv3 condition (**Figure 3.2e**). At the same target site, a series of 5mer deletions were encoded in the pegRNA, ranging from 5 to 25 bp in length. Editing frequencies were 1.9% and 1% for SaPE and sSaPEv3, respectively, for a 5 bp deletion (**Figure 3.2f**). These rates decreased as the deletion length increased. 25 bp deletions were still detected, albeit at a lower frequency than other deletion lengths.

We repeated some of the above experiments in U2-OS cells to explore cell type dependency. Overall, no trend emerged in regard to predicting optimal editing frequency at a given target locus, edit type, and pegRNA design. Editing efficiencies were in the single digits across all loci assayed, although the *HEK3* locus offers promise moving forward.

Altering the dimerization domain

Following extensive characterization of the editing frequencies, we aimed to better understand the split protein recombination requirements. To visualize reassociation of split SaPE, we replaced the Npu split intein with split NanoLuc (**Figure 3.3a**). Split NanoLuc (sNanoLuc, sNL), with a K_D of 700 pM, is comprised of an 18 kDa N-terminal fragment, LgBiT, and a 13 amino acid C-terminal portion, HiBiT[168]. When co-transfecting the two halves of sSaPE-sNanoLuc, we saw extensive nuclear reconstitution of NanoLuc in a vast majority of cells using

bioluminescence microscopy (**Figure 3.3b**). Similar editing rates to sSaPEv3 were also seen, indicating that sNanoLuc is sufficient to act as a dimerization domain in the context of SaPE (**Figure 3.3c**). The lack of requiring covalent association of the two halves of SaPE was further evaluated by utilizing a catalytically inactive version of the Npu intein (C1A). In these constructs, the two intein components can associate but not self-splice. No covalent full-length SaPE is formed, yet editing rates are on par with the catalytically active intein version (**Figure 3.3c and d**). This data further validated that covalent association of the two halves of split SaPE is not required to reconstitute active SaPE. The use of sNanoLuc also offers a convenient system in which visual confirmation of recombined SaPE is possible, an approach which can be utilized with AAV delivery.

Packaging and delivery of SaPE in AAV

Next, we cloned the N- and C-terminal portions of sSaPE versions 1 and 3 into AAV genomes containing flanking ITRs. For N-terminal sSaPE, two orientations of the U6 promoter-pegRNA cassette were tested, either in tandem or in reverse alignment to the protein expression cassette (**Figure 3.4a**). The C-sSaPEv3 AAV genome is 4.8 kb in length while the N-terminal genomes are approximately 3.5 kb. Before packaging into virus, AAV plasmids were co-transfected targeting *GFP(L202S)* to ensure editing still occurred in the different context. Indeed, editing rates were in line with previous plasmid designs, even with a co-transfection as opposed to a triple transfection (**Figure 3.4b**). Next, AAV-DJ

was packaged and titered using qPCR, yielding ~$2x10^{10}$ viral genome copies (g.c.)/µl. A control transduction with tdTomato as the encoded genome exhibited robust delivery and fluorescent protein expression in HEK-293T cells (**Supplementary Figure 3.4**), ensuring proper production and purification of the virus. We co-transduced sSaPE AAVs in *GFP(L202S)* HEK-293T cells and first assessed protein recombination. At both three and five days post-transduction, full-length recombined protein was evident via western blot (**Figure 3.4c**). Assessing GFP fluorescence restoration in *GFP(L202S)*, we see a concentration dependent increase in editing (**Figure 3.4d**). While the GFP restoration frequency is low, co-delivery of sSaPEv3-sNanoLuc AAVs in these cells also showed a low number of cells expressing recombined protein, potentially due to a low co-transduction efficiency (**Supplementary Figure 3.4**). However, co-transduction in U2-OS cells resulted in an order of magnitude increase in delivery and recombination efficiency as visualized by bioluminescence microscopy (**Figure 3.4e**). Taken together, we have begun the proof of concept work necessary to allow prime editor to be delivered via dual AAVs for potential future clinical applications.

Discussion

We have demonstrated that an *S. aureus* prime editor presents a platform upon which in vivo prime editing can occur. While the editing rates we achieved in vitro were overall low (typically 0.5 to 1%) and not likely to be effective for many diseases, this approach offers a starting point to further refine and improve. It is important to note that comparably poor prime editing rates, albeit in plants and protoplasts, were seen in other published reports with SpPE[160,167]. Multiple engineering approaches can be undertaken to increase the efficiency of the system. To enhance activity with SaCas9, directed evolution of MMLV-RT could be employed to increase activity on pegRNA:R-loop DNA in a method akin to a recently evolved adenosine base editor[169]. A circularly permuted version of SaCas9 to facilitate the optimal RT fusion location could also be created, similar to what was also performed with base editors[170], to increase access to the R-loop DNA. Another reason SaCas9 might have diminished prime editing rates could be attributed to a decreased residency time on the DNA target[159]. Chemically or genetically disrupting factors involved in removing Cas9 from genomic DNA has been successful at increasing other Cas9-fused effector functions. In one such study, genetic knockdown of the histone chaperone FACT increased Cas9 residence on DNA and led to improved epigenetic marking and CRISPRi[171]. It is also interesting to note that using the sSaPE-sNanoLuc platform, we see that the construct is being expressed in a majority of the cells (**Figure 3.3b**). A recent study examining pegRNA design principles does provide insight into optimal pegRNA

design, although how well it translates to SaPE remains to be seen[172]. Further investigation, aided by the preceding suggestions, is warranted into studying the underlying reasons behind the disconnect between expression and prime editing.

A downside of SaCas9 is the more limited reach in genomic space with a longer PAM sequence of NNGRRT. This limitation is especially relevant as mutations further away from the +1 site tended to have lower editing rates. However, iterations such as SaCas9 (KKH) have lessened the PAM specificity (PAM = NNNRRT)[173]. Such codon changes could feasibly be incorporated to SaPE to expand the targeting range. Additionally, SaPE is a less ideal system for a PE3 type system, wherein a second gRNA nicks the non-editing strand to encourage the desired editing outcome, due to the stricter PAM specificity. A less stringent PAM could aid in a broader exploration of PE3 with SaPE.

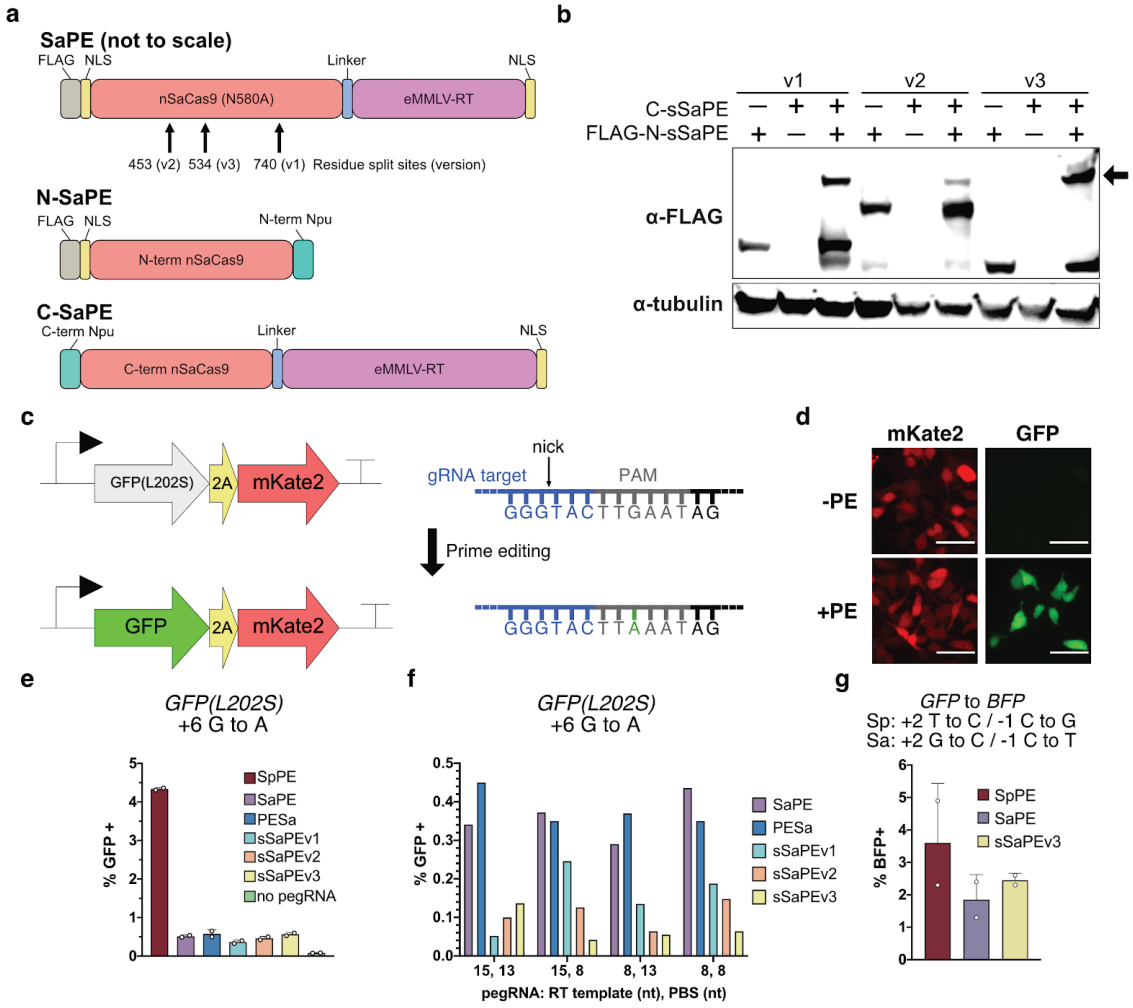
Once we are able to demonstrate in vitro AAV delivery of sSaPE, the next phase will be to test in vivo. The ultimate goal is to perform delivery using targeted delivery approaches such as HUH-AAV[78]. This could have the advantage of decreasing the necessary administered dose, potentially limiting immunogenicity, and alleviating non-target cell editing concerns. In conclusion, this work provides a baseline platform for prime editor to be delivered in vivo. Through making an *S. aureus* prime editor and combining with trans splicing intein technology, we are able to package prime editor into dual AAVs for delivery. The anticipated engineering advances that will be made with prime editing components, such as increased DNA residency time or enhanced RTs, can readily be incorporated into

this platform to increase editing efficiencies and move prime editing towards the clinic.

Acknowledgements

We would like to acknowledge the core facilities at the University of Minnesota who assisted with this work: University Imaging Center (UIC), University Flow Resource Center (UFRC), and the Viral Vector and Cloning Core (VVCC).

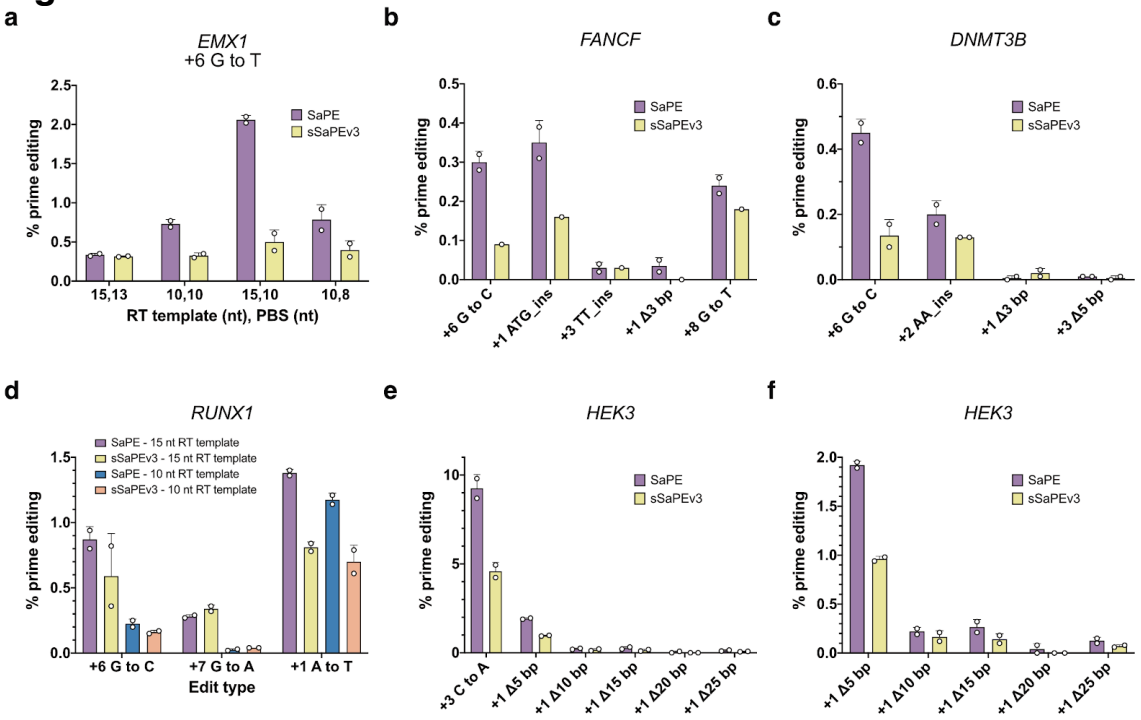
Figure 3.1



Enabling precise genome modifications using a split *S. aureus* prime editor. (**a**) Schematic of both full-length *S. aureus* prime editor (SaPE) and split SaPE (sSaPE). (**b**) Western blot of plasmid-based expression of three different versions of sSaPE. Successful trans intein splicing is denoted by the arrow next to the higher molecular weight band. (**c**) Model system for analyzing editing by SaPE using a GFP(L202S) stable reporter line to restore GFP fluorescence. A single point mutation (G to A) that lies within SaPE's PAM is required. The antisense DNA sequence is shown. (**d**) Representative fluorescence microscopy images of

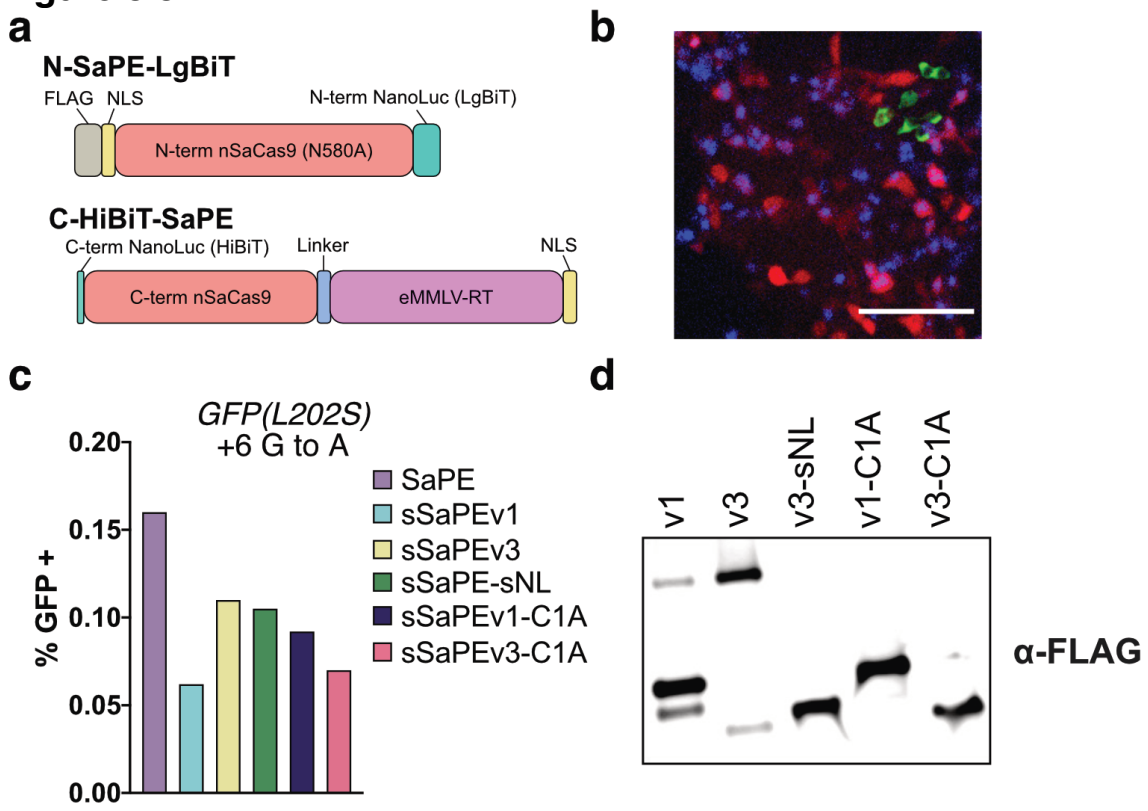
unedited (-PE) and edited (+PE) HEK-293T stably expressing the GFP(L202S)-2A-mKate2 reporter. Scale bar = 20 μm. (**e**) Quantitation of prime editing in HEK-293T cells using flow cytometry. *S. pyogenes* prime editor (SpPE) is compared to full-length SaPE, N-terminally fused RT-nSaCas9 (PESa), and three versions (v1-v3) of sSaPE. No pegRNA corresponds to transfection of SaPE only. Data are representative of multiple independent experiments. (**f**) Targeting of *GFP(L202S)* reversion with pegRNAs of differing 3' extension lengths (PBS = primer binding site). (**g**) GFP to BFP editing in HEK-293 cells stably expressing wildtype GFP. Individual data points represent biological replicates.

Figure 3.2



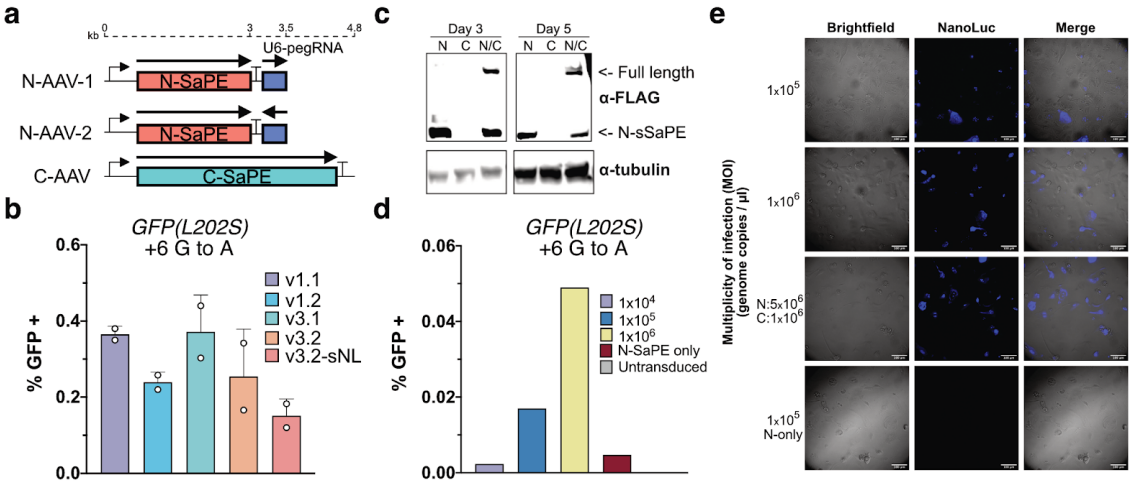
SaPE is capable of diverse edit types across various genomic loci. (**a-f**) Next-generation sequencing readouts of targeting of SaPE and sSaPEv3 to indicated loci (**a**) *EMX1*, (**b**) *FANCF*, (**c**) *DNMT3B*, (**d**) *RUNX1*, and (**e,f**) *HEK3* with differing edit types and pegRNA architectures. (**f**) contains the same 5mer deletion series data as (**e**), but with a scaled y-axis.

Figure 3.3



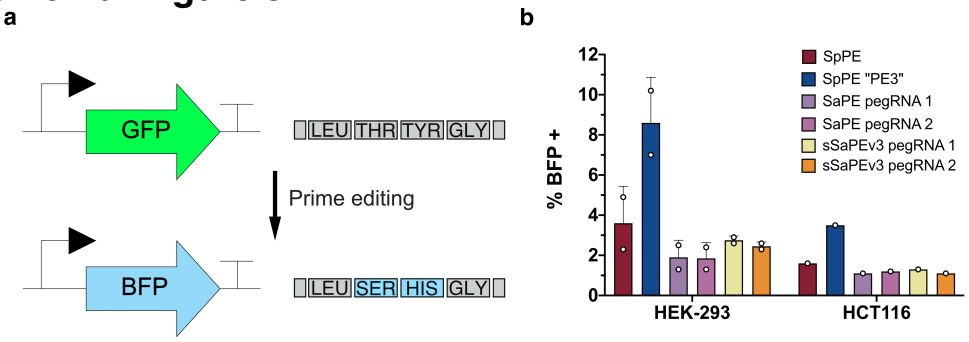
Split NanoLuc substitutes as an effective dimerization domain and visualization tool. (**a**) Diagram of sSaPE-split NanoLuc (sNL) (not to scale). (**b**) Composite image of stable GFP(L202S)-2A-mKate2 HEK-293T cells transfected with sSaPE-sNL. Colors are as follows: red = mKate2; blue = NanoLuc; green = GFP. Scale bar = 120 μm. (**c**) Prime editing rates at *GFP(L202S)* locus in HEK-293T cells. (**d**) Western blot of co-transfection of indicated versions of split SaPE. Only the N-terminal fragment is FLAG labeled. The top band corresponds to covalently recombined SaPE.

Figure 3.4



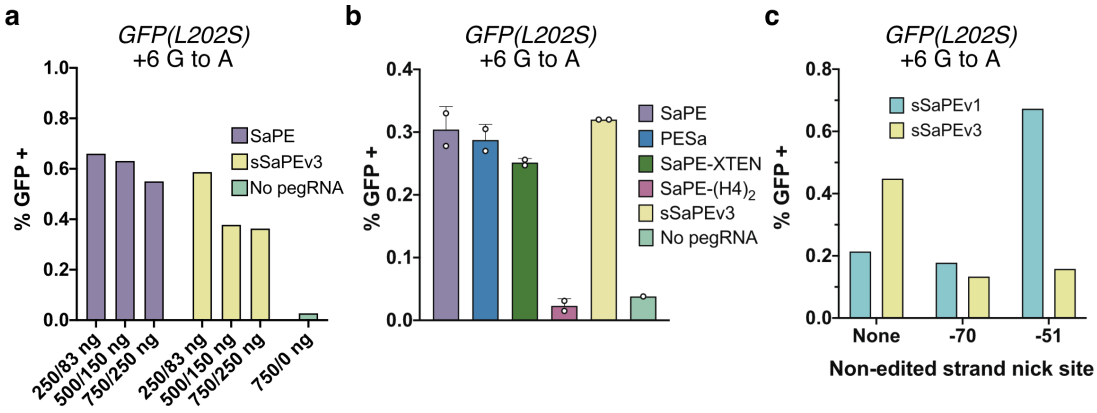
Packaging of split SaPE into AAV. (**a**) Scheme of different layouts of sSaPE in AAV genome with indicated sizes of expression cassettes. (**b**) Plasmid-based prime editing of SaPE encoded in AAV plasmids. Version 3.2-sNL corresponds to sNanoLuc replacing Npu intein. (**c**) Western blot of HEK-293T cells transduced with the indicated virus(es) for 3 or 5 days. Only the N-terminal fragment is FLAG labeled. (**d**) Co-transduction of AAVs expressing sSaPEv3 at the indicated multiplicity of infections (genome copies / cell) targeting *GFP(L202S)*. (**e**) Bioluminescence microscopy images of co-transduced AAV-sSaPEv3-sNanoLuc in U2-OS cells after 4 days.

Supplemental Figure 3.1



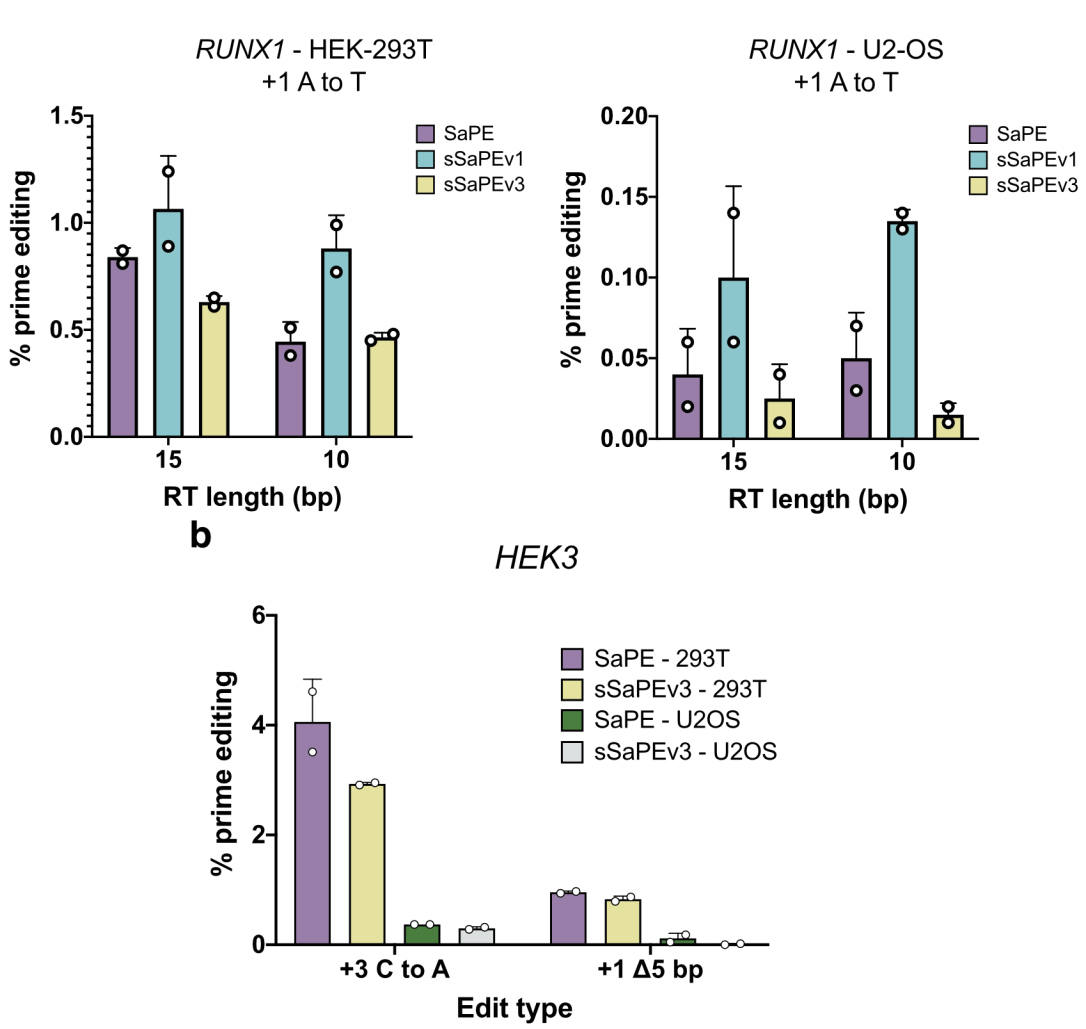
Conversion of GFP to BFP. (**a**) Editing scheme of converting GFP to BFP through two point mutations. (**b**) Percentage of GFP converted to BFP in HEK-293 and HCT-116 cells electroporated with the indicated prime editors as analyzed by flow cytometry. Two different pegRNAs were trialed for SaPE.

Supplemental Figure 3.2



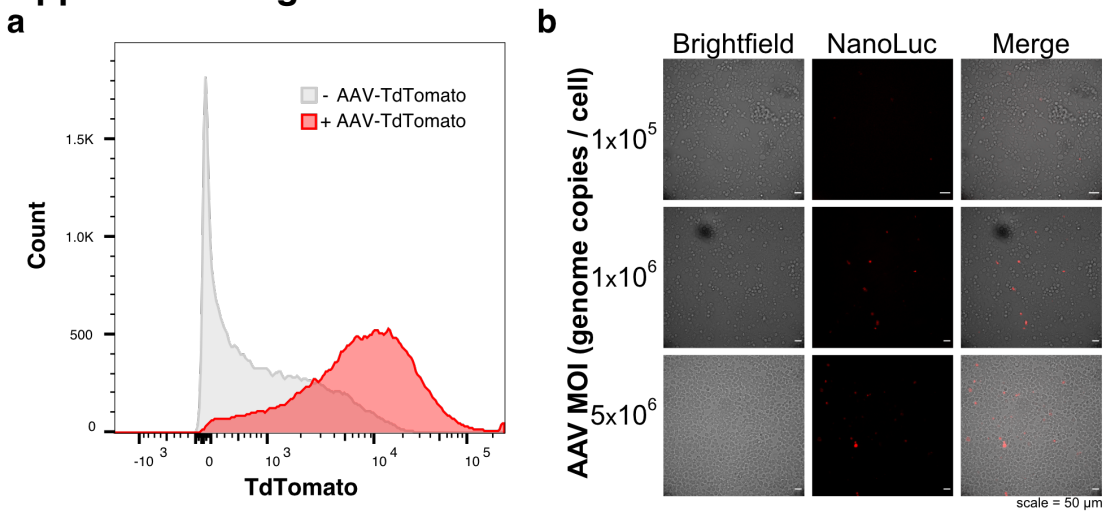
Optimization of SaPE using GFP(L202S) reporter system. (**a**) Titration of both full-length and split SaPE. DNA concentrations on the x-axis correspond to prime editor/pegRNA amounts. (**b**) Comparison of various versions of SaPE, including altering the linker between nSaCas9 and RT. (**c**) Combining SaPE with a gRNA nicking on the non-edited strand at the indicated position to drive incorporation of the desired edit.

Supplemental Figure 3.3



Prime editing in U2-OS cells. (**a**) Next generation sequencing results of prime editing at the *RUNX1* locus of a +1 A to T mutation in both HEK-293T (left) and U2-OS cells (right). Different RT template lengths in the pegRNA are used while the PBS is held constant at 8 bp in length. (**b**) Prime editing comparison at the *HEK3* locus in different cell types.

Supplemental Figure 3.4



AAV-mediated delivery of tdTomato and SaPE. (**a**) AAVs containing TdTomato as the cargo were transduced in 293T-GFP(L202S)-2A-mKate2 reporter cell line. 48 hr post-transduction, cells were analyzed using flow cytometry and compared to non-transduced cells. (**b**) Microscopic images of co-transduction of sSaPEv3-sNanoLuc encoded in AAVs in HEK-293T cells. Scale bar is 50 μm. MOI = Multiplicity of Infection.

Supplemental Table 3.1

pegRNA sequences.

Target gene	Spacer sequence (5' to 3')	Modification	RT template	PBS
EMX1	GGCCTCCCCAAAGCCTGGCCA	+6 G to T	ctggccactAcctgg	ccaggctttgggg
EMX1	GGCCTCCCCAAAGCCTGGCCA	+6 G to T	cactAcctgg	ccaggctttg
EMX1	GGCCTCCCCAAAGCCTGGCCA	+6 G to T	ctggccactAcctgg	ccaggctttg
EMX1	GGCCTCCCCAAAGCCTGGCCA	+6 G to T	cactAcctgg	ccaggctt
L202S-GFP	TTGCTCAGGGCCGACTGGGTAC	+6 G to A	ACCACTAttTaAGtA	CCCAGTCgGCCCT
L202S-GFP	TTGCTCAGGGCCGACTGGGTAC	+6 G to A	ACCACTAttTaAGtA	CCCAGTCg
L202S-GFP	TTGCTCAGGGCCGACTGGGTAC	+6 G to A	ttTaAGtA	CCCAGTCgGCCCT
L202S-GFP	TTGCTCAGGGCCGACTGGGTAC	+6 G to A	ttTaAGtA	CCCAGTCg
RUNX1	GTACTCACCTCTCATGAAGCACT	+6 G to C	CTTCGTACCgACAGT	GCTTCATG
RUNX1	GTACTCACCTCTCATGAAGCACT	+6 G to C	TACCgACAGT	GCTTCATG
RUNX1	GTACTCACCTCTCATGAAGCACT	+7 G to A	CTTCGTACtCACAGT	GCTTCATG
RUNX1	GTACTCACCTCTCATGAAGCACT	+7 G to A	TACtCACAGT	GCTTCATG
RUNX1	GTACTCACCTCTCATGAAGCACT	+1 A to T	CTTCGTACCCACAGa	GCTTCATG
RUNX1	GTACTCACCTCTCATGAAGCACT	+1 A to T	TACCCACAGa	GCTTCATG
FANCF	gtagggccttcgcgcacctca	+1 A to C	gaagggattccatgC	ggtgcgcg
FANCF	gtagggccttcgcgcacctca	+1 A to C	gattccatgC	ggtgcgcg
FANCF	gtagggccttcgcgcacctca	+2 G to C	gaagggattccatCa	ggtgcgcg
FANCF	gtagggccttcgcgcacctca	+2 G to C	gattccatCa	ggtgcgcg
FANCF	gtagggccttcgcgcacctca	+3 T to G	gaagggattccaGga	ggtgcgcg
FANCF	gtagggccttcgcgcacctca	+3 T to G	gattccaGga	ggtgcgcg
GFP/BFP	GAAGCACTGCACGCCGTAGGT	+2 G to C / -1 C to T	GACCACCCTGAgCc	ACGGCGTGCAGT
GFP/BFP	GAAGCACTGCACGCCGTAGGT	+2 G to C / -1 C to T	ACCCTGAgCc	ACGGCGTG
HEK3	tctgcttctccagccctggc	+3 C to A	ttgacccagTcc	agggctggag
HEK3	tctgcttctccagccctggc	+1 5mer deletion	ggattgaccc	agggctggag
HEK3	tctgcttctccagccctggc	+1 10mer deletion	cccaaggatt	agggctggag
HEK3	tctgcttctccagccctggc	+1 15mer deletion	tgggccccaa	agggctggag
HEK3	tctgcttctccagccctggc	+1 20mer deletion	cagtctgggc	agggctggag
HEK3	tctgcttctccagccctggc	+1 25mer deletion	gtgctcagtc	agggctggag
FANCF	gatgttccaatcagtacgca	+6 G to C	cggcgactGtctgc	gtactgattgga
FANCF	gatgttccaatcagtacgca	+1 ATG_ins	cggcgactctctgcCAT	gtactgattgga
FANCF	gatgttccaatcagtacgca	+3 TT_ins	cggcgactctcAAtgc	gtactgattgga

FANCF	gatgttccaatcagtacgca	+1 tgc del	agacggcgactctc	gtactgattgga
FANCF	gatgttccaatcagtacgca	+8 G to T	cggcgaAtctctgc	gtactgattgga
DNMT3B	ggtggcactgcggctggaggt	+6 G to C	ctttaaccGccacc	tccagccgcagt
DNMT3B	ggtggcactgcggctggaggt	+2 AA ins	ttaacccccacTTc	tccagccgcagt
DNMT3B	ggtggcactgcggctggaggt	+1 ggt del	tttaaccccc	tccagccgcagt
DNMT3B	ggtggcactgcggctggaggt	+3 G_5_del	ctccgctttaaacc	tccagccgcagt

For spacer sequences that don't begin with a G nucleotide, one was added to aid in U6-driven expression of the pegRNA. Modifications are listed relative to the +1 nick site 3 nucleotides upstream from the PAM on the sense DNA strand.

Supplemental Table 3.2

Primers used for genomic DNA amplification.

Name	Sequence (5' to 3')
SpEMX1_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT CAGCTCAGCCTGAGTGTTGA
SpEMX1_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT CTCGTGGGTTTGTGGTTGC
L202S_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT TTCAAGATCCGCCACAACAT
L202S_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT GTATAGTTCATCCATGCCGAG
SaEMX1_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT gcaaccacaaacccacgag
SaEMX1_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT agacacggagagcagctg
RUNX1_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT GAGGGTGCATTTTCAGGAGG
RUNX1_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT CAAGCTGCCATTTCATTACAGG
FANCF_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT ccagagtcaaggaacacgga
FANCF_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT acgtaggtagtgcttgagacc
DNMT3B_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT gaacccaggtagccagagac
DNMT3B_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT tcctttcaacccgaacggag
HEK3_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT ggaaacgcccatgcaattag
HEK3_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT cccagccaaacttgtcaacc

Supplemental Table 3.3

SaPE Protein sequences.

Protein	Sequence (N-to-C)
SaPE (SaCas9 H840A - linker - eMMLV-RT)	MKRTADGSEFESPKKKRKVKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFK EANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSK ALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILV NEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSE DIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFN RLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEG KCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRT PFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEY KEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLN GLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEE TGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRF DVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNN DLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSI KKYSTDILGNLYEVKSKKHPQIIKKGSGGSSGGSSGSETPGTSESATPESSGGS SGGSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKK PGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFF CLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLAD FRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQK QVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGF AEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVD EKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVV ALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQ EGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDS RYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQK GHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPK KKRKV*
FLAG-N-split SaPE-N-Npu version 1	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAKRNYILGLDIGITS VGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVN EVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDY VKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEW YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQII ENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKE IIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLK AINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFI QSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTG

	KENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSF NNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKE YLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSING GFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAECLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQW HDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLP N*
C-Npu-C-SaPE version 1	MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASNSMPEIETEQEYKEIFITPHQ IKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDND KLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKY SKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGV YKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGEL YRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDIL GNLYEVKSKKHPQIIKKGSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTL NIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATS TPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYR PVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPT SQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPD LILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAP LYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYA KGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMG QPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRK AGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFA TAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAE ARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV*
FLAG-N-split SaPE-N-Npu version 3	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAKRNYILGLDIGITS VGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVN EVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDY VKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEW YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQII ENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKE IIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLK AINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFI QSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTG KENAKYLIEKIKLHDMQEGKCLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNN GNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFE RELDLMRVDNLPN*
C-Npu-C-SaPE version 3	MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASNCLYSLEAIPLEDLLNNPFNY EVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHI LNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSY FRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKE WKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKY SHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPE KLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIK

	KIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLD VIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDL LNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKK HPQIIKKGSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEYRLHE TSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPM SQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWR DPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLL LAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWL TEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLF NWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQH NCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEV IWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRR GWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQA ARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV*
N-term split NanoLuc	EDFVGDWEQTAAYNLDQVLEQGGVSSLLQNLAVSVTPIQRIVRSGENALKIDIH VIIPYEGLSADQMAQIEEVFKVVYPVDDHHFKVILPYGTLVIDGVTPNMLNYFGR PYEGIAVFDGKKITVTGTLWNGNKIIDERLITPDGSMLFRVTINS
C-term split NanoLuc	VSGWRLFKKIS

Supplemental Note 3.1

Cloning SaPE pegRNAs using golden gate assembly.

The pegRNAs were cloned using a protocol adapted from the Liu lab (Anzalone, *et al.*, 2019). These modifications were made to incorporate the *S. aureus* gRNA scaffold sequence and to add some streamlined features in regard to vector digestion and golden gate assembly cycling conditions.

Step 0: Design of oligonucleotides

Protospacer oligonucleotides
Forward: CACC ...(spacer sequence).... GTTTT
Reverse: TACTAAAAC …(reverse complement spacer)...
Note: Add a G prior to spacer sequence if it doesn't begin with a G (and corresponding C on reverse oligonucleotide).

pegRNA 3' extension oligonucleotides
Forward: GAGA …(RT template + PBS)...
Reverse: AAAA …(reverse complement)...

S. aureus ***phosphorylated*** *scaffold oligonucleotides*
Forward:
/5Phos/agtactctggaaacagaatctactaaaacaaggcaaaatgccgtgtttatctcgtcaacttgttggc

Reverse:
/5Phos/tctcgccaacaagttgacgagataaacacggcattttgccttgttttagtagattctgtttccagag

Step 1: Anneal oligonucleotides

In IDT annealing buffer (30 mM HEPES pH 7.5, 100 mM KAc), add equimolar amounts of forward and reverse oligonucleotides to 1 µM. Heat to 95 °C for 3 minutes then cool to room temperature at 0.1°C/s. Unused annealed oligonucleotide can be stored at -20°C.

Step 2: Golden gate assembly reaction

Undigested pU6-pegRNA-RFP acceptor (Addgene #132777)	1 µL @ 250 ng/µL
Annealed protospacer oligonucleotides	1 µL @ 1 µM
Annealed pegRNA 3'-extension oligonucleotides	1 µL @ 1 µM
Annealed **phosphorylated** sgRNA scaffold oligonucleotides	1 µL @ 1 µM
BsaI-HFv2 (NEB)	0.25 µL
T4 DNA ligase (NEB)	0.50 µL
10x T4 DNA ligase buffer (NEB)	1 µL

H_2O	4.25 µL
Total reaction volume	**10 µL**

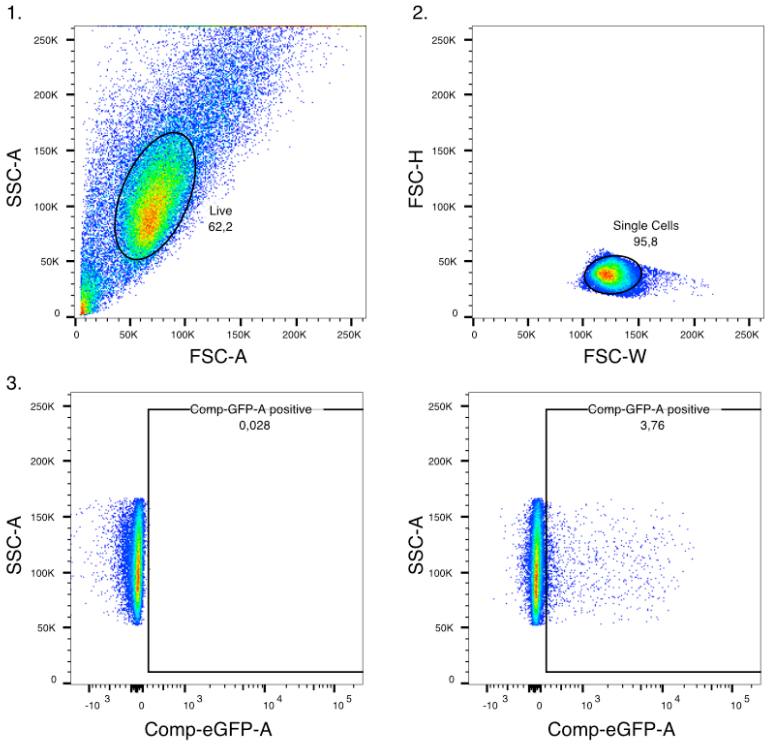
Perform the following program in a thermocycler:
Cycle 10x:
5 min at 37 °C
10 min at 16 °C
Following cycles, incubate at 5 min at 55°C followed by 5 min at 85°C then hold at 12°C

Step 3: Transformation
Transform 1 µL of the assembly reaction into competent *E. coli* cells of your choosing. Plate on LB Agar, incubate overnight at 37 °C, and inoculate **non-red** colonies for DNA miniprep.

Supplemental Note 3.2

Flow cytometry gating strategy and representative plots.

1. Gating on live cells SSC-A vs FSC-A
2. Gating on single cells (omitting aggregates) FSC-H vs FSC-W
3. Gating on eGFP negative vs positive (negative control compared to positive edit)



Chapter 4

In vivo gene delivery to prostate tumors via an anti-fibroblast activation protein antibody-AAV composite

This chapter is intended as a preliminary start for a future manuscript and as a reference for future project researchers

Zdechlik, A.C., Hintz, H.M., LeBeau, A.M., Schmidt, D.A.

All authors jointly conceived idea and designed experiments. A.C.Z. and H.M.H. carried out experiments together. H.M.H. analyzed data and created figures. A.C.Z. authored the manuscript.

Preface

Prostate cancer is a heterogeneous disease usually diagnosed in late stages, making treatment challenging and frequently ineffective[174]. There is a need for new treatment and diagnostic methods. Fibroblast activation protein alpha (FAP) is a surface marker expressed on stromal cancer-associated fibroblasts that has recently been shown to be an effective marker of prostate tumors[175]. Our lab has described a technology to redirect adeno-associated virus (AAV) infection to be dependent upon an antibody-antigen interaction[78]. In this study, we aimed to create a cancer gene therapy vector by conjugating an anti-FAP antibody to the capsid of this modified AAV. We characterized formation of this composite using electron microscopy. Initially, we delivered a near-infrared fluorescent protein to mice with FAP+ xenografted tumors. Future studies will deliver therapeutic genes to treat the tumors by inhibiting growth.

Introduction

Prostate cancer is a leading cause of cancer death in men, responsible for 375,304 deaths in 2020 alone [176]. Treatment options are typically limited to surgical removal of the prostate and/or castration, hormone therapy, and radiotherapy, all of which have major side effects [174]. Patients with aggressive metastatic prostate cancer frequently stop responding to standard of care treatments and are left with few options [177]. There is a need for new therapeutics capable of treating heterogeneous, late-stage prostate tumors.

Recently, an antibody that binds murine fibroblast activation protein alpha (FAP), a protein expressed in stromal fibroblasts in prostate and other cancers has been characterized [178]. The stroma can account for up to 90% of tumor mass, making it an attractive target for chemotherapy [179]. FAP expression is largely restricted to stromal cancer associated fibroblasts (CAFs), meaning that any therapeutic delivered to FAP-expressing cells would minimally effect healthy cells [180]. Past studies have demonstrated that killing FAP+ stromal fibroblasts inhibits growth of xenografted tumors, further supporting the idea that targeting the CAFs is an effective treatment[180]. Additionally, in in metastatic castration-resistant prostate cancer, FAP has been shown to be a reliable tumor marker, providing an improved therapeutic and diagnostic target [175].

Targeted chemotherapeutics are key to minimizing side-effects and off-target cell death. One way to achieve targeted delivery of therapeutics is with antibody-

drug conjugates. One antibody-drug conjugate (ADC) that has been used is composed of an anti-tumor endothelial marker 8 (TEM8) antibody conjugated to monomethyl auristatin E (MMAE), a drug commonly used in ADCs [181,182]. The ADC effectively homed to TEM8, which is expressed in CAFs, but the drug has limited use because MMAE response varies greatly between tumor cell lines [182]. With any chemotherapeutic, it is likely that the carcinoma will eventually grow resistant to treatment [183]. While there may be alternate routes of treatment, the limited plasticity of small molecules for ADCs makes them a cumbersome route of action.

In contrast to small molecule chemotherapeutics, cancer gene therapy is highly plastic. Gene therapy for prostate cancer has gained popularity in recent years, with several therapeutics progressing to phase I and II clinical trials [184]. The genes delivered in these strategies broadly either recruit the immune system or attempt to induce cell death. Many different therapeutic gene products (i.e. siRNA, RNAi, proteins, miRNA, etc.) and their mechanisms of action have been characterized, [185]. Additionally, once a delivery mechanism has been selected, modifying treatment from one gene to another is relatively easy and quick. Additionally, even small vectors typically have space to deliver more than one gene product, making them an ideal platform for simultaneous delivery of a therapeutic and diagnostic. One gene delivery vector that is gaining popularity for cancer gene therapy is adeno-associated virus (AAV) [71,72]. However, natural serotypes of AAV do not specifically target cancerous cells, resulting in delivery of therapeutic genes

to healthy tissue. The main challenge precluding cancer gene therapy from widespread adoption at this point is the lack of a safe, targeted delivery vector [74].

Recently, our lab has characterized a system of covalently conjugating AAV to an antibody to facilitate antibody-mediated infection [78]. The vector is modified via engineering one of the AAV capsid proteins to contain a small single-stranded DNA (ssDNA) binding domain called an HUH tag. In this context, the HUH tag forms a covalent bond with its matched sequence of ssDNA [120]. An antibody can be attached to this ssDNA sequence using commercially-available copper-free click chemistry. By reacting this ssDNA-antibody with the HUH-modified AAV capsid, a fully covalent antibody-AAV composite is formed. Additionally, the native tropism of the virus is removed by mutating regions of the capsid responsible for heparin sulfate and beta-integrin binding [34]. Resultantly, this modified virus is only able to bind and infect cells via the antibody-antigen interaction.

To further characterized this antibody-AAV composite, we imaged capsids using transmission electron microscopy. It is known that the HUH tag used in the capsid is not 100% efficient, and that rates of incorporation of modified viral proteins can vary[120]. Viral titer is calculated by number of genome-containing capsids, whether they are conjugated to an antibody or not. Given this information, we wanted to know approximately what percentage of genome-containing capsids were conjugated to an antibody. This measure should provide a more accurate “effective titer” of the number of viruses capable of infecting cells via the FAP-

antibody interaction, which would be a truer point of comparison of infectivity relative to unmodified AAV-DJ.

We aimed to selectively deliver genes to CAFs by conjugating the anti-FAP antibody to this modified AAV capsid. Initially, we delivered a fluorescent protein (miRFP670) to confirm specific delivery of the gene to the tumor. miRFP670 is a near-infrared fluorescent protein that can be imaged *in vivo*[186]. Data from two experiments were collected with two or three biological replicates for each experimental group. We also collected electron microscopy data characterizing the conjugated virus to gain a better idea of conjugation efficiency. Future work will investigate whether tumor growth can be inhibited by delivering therapeutic genes alongside imaging components.

Methods & Materials

Antibody Production and Conjugation

Anti-FAP antibody was produced as previously described [178]. Briefly, both domains were cloned into IgG expression vectors and cotransfected into HEK293 cells. Antibody was harvested from serum and purified using a HiTrap Protein A HP column (GE Healthcare). Human IgG isotype control was purchased from ThermoFisher. 3' amino-labeled mMobA *ori* ssDNA sequence was ordered from Integrated DNA Technologies (Coralville, IA). Each antibody and ssDNA were conjugated using an abcam Oligonucleotide Conjugation Kit at an estimated ratio of 3:1 ssDNA:antibody. Free ssDNA was removed and conjugate formation was verified via SDS-PAGE.

Virus Production and Conjugation

All viruses used in this study were generated by the University of Minnesota Viral Vector and Cloning Core except for virus used in electron microscopy. The experimental viral capsid AAV-DJ VP2-mMobA was prepared as previously described[78]. All viral genomes encoded miRFP670 under control of CAG promoter. The coding sequence for miRFP670 was a gift from Vladislav Verkhusha (Addgene plasmid # 79987 ; http://n2t.net/addgene:79987 ; RRID:Addgene_79987) [186]. Immediately before injection, each virus was conjugated to appropriate antibody-ssDNA. Virus was incubated at 37C for 45 minutes with 2mM MgCl2, 2 mM MnCl2, 1:20 salmon sperm DNA, and excess antibody-ssDNA in PBS. After incubation,

each conjugation reaction was diluted to working volume in PBS and filtered through a 0.2-micron syringe filter. Conjugated viruses were kept on ice until injection into animals.

Immunoelectron Microscopy

Virus for electron microscopy was prepared by Vigene Biosciences from plasmid DNA encoding AAV-DJ VP2-mMobA, AAV-DJ VP2/3ΔHBD, and pAAV-HSV-TK. Virus was purified using IDX gradient ultracentrifugation to remove cellular contaminants and empty capsids. Prior to imaging, virus was reacted with a mouse CD3 antibody conjugated to the mMobA ssDNA sequence in HUH reaction conditions as described. The reaction was then incubated with an anti-mouse 10 nm immunogold secondary antibody (Millipore Sigma) for 30 minutes. As a negative control, virus not conjugated to a primary antibody was incubated with the immunogold secondary antibody and imaged (data not shown). Transmission electron microscopy was performed by Wei Zhang in the University of Minnesota Characterization Facility. Antibody-virus was applied to charged carbon grids and negative stained with uranylformate. Samples were imaged using an FEI Tecnai T12 Transmission Electron Microscope.

Animal Models

All animal work was in accordance with a UMN Institutional Animal Care and Use Committee protocol. Xenografts were generated as previously described [178].

Briefly, athymic nude mice (Envigo) were subcutaneously injected with R1-EnzR^{FAP} cells, an immortalized prostate cancer cell line (CWR-R1-enzalutamide resistant/luciferase^{+}) lentivirally transduced to express human FAP. Metastatic colony growth was monitored over 21 days leading up to virus injection. Antibody-AAVs and appropriate controls were administered at a total volume of 100uL via tail vein injection.

In Vivo Near-Infrared Imaging

miRFP670 and luciferase were imaged using an IVIS Spectrum In Vivo Imaging system with assistance of Guillermo Marques at the University of Minnesota- University Imaging Centers. Ten minutes prior to imaging, all animals were injected with D-luciferin. Images were collected 24-, 48-, 72-, 144-, 168-, and 240-hours post-infection. After final imaging, animals were sacrificed. Liver, tumor, and blood from each animal was collected and flash-frozen for future analysis.

Results

αFAP-specific Infection

In the first experiment, all mice were infected with the αFAP-AAV infected at a multiplicity of infection (MOI) of approximately $3x10^{11}$ genome copies, which corresponds to approximately $1.5x10^{13}$ GC/kg, roughly one order of magnitude less than the standard AAV dosage for mice[187]. Mice were imaged seven days post-infection and harvested organs were imaged for miRFP expression immediately following sacrifice and harvest. Fluorescence was observed in the eyes of mice infected with αFAP-AAV, no antibody, and AAV-DJ. However, no fluorescence was observed in the eyes post-dissection (**Figure 4.2b**). Surprisingly, no fluorescence was visible in the liver of mice infected with AAV-DJ or αFAP-AAV (**Figure 4.2c**). In contrast to *in vivo* imaging, dissected tumor fluorescence was similar between AAV-DJ and αFAP-AAV (**Figure 4.2c**). In response to these data, we were interested in capturing how viral genome expression changed over time and determining whether αFAP-AAV would be effective at lower doses.

In the second experiment, three MOIs of $2x10^{11}$(MOI 1), $2x10^{10}$(MOI 2), and $2x10^{9}$(MOI 3) were tested for the experimental αFAP-AAV group. All other viruses were administered at an MOI of $2x10^{11}$ GC/animal ($8.6x10^{12}$ GC/kg). A time course of images at 48, 72, 144, 168, and 244 hours post-injection was collected to capture peak expression of miRFP670; expression was highest at the final time point when mouse tumors had grown large enough to necessitate sacrifice (**Figure 4.3a**). It is possible that expression would have been higher after this time point. As a control, the same base capsid conjugated to αFAP was conjugated to a non-

targeting IgG and administered at the highest dose (MOI 1). This group was added to test the idea that an interaction between an antibody and the tumor, not the antibody-antigen binding was facilitating infection. There was significant variability observed between animals in the same groups, particularly between the αFAP-AAV-infected animals (**Figure 4.3b**). Animals infected with the IgG-AAV had higher expression of miRFP670 in the tumor than those infected with either AAV-DJ or the HUH-AAV without an antibody. However, αFAP-AAV provided the highest level of expression on the 10^{th} day after injection (**Figure 4.3c**). The dynamics of miRFP670 expression were similar between treatment groups, with fluorescence slowly increasing over time and peaking at 240 hours for IgG-AAV and αFAP-AAV groups (**Figure 4.3a**).

Immunoelectron microscopy

Electron microscopy of the capsid revealed a very low percentage of capsids (3.8%) are conjugated to an antibody (**Figure 4.4**). Very few empty capsids were observed, meaning that capsid modification did not interfere with IDX gradient purification. Some gold particles were observed close to viral capsids in the negative control, suggesting that there is some rate of random association between the two. Further experiments will be necessary to calculate conjugation efficiency in optimized experimental conditions.

Discussions and Future Directions

Initially, it appeared that addition of the anti-FAP antibody to the capsid of AAV strongly directed infection to the tumor via the FAP-antibody interaction. In the first set of experiments, the αFAP-AAV strikingly infected the tumor more efficiently than either the negative or positive control (**Figure 4.2**). It is unclear why there was no fluorescence observed in the liver of mice infected with AAV-DJ. AAV-DJ was selected specifically to be highly infective to the liver and evade preexisting antibodies [39]. The reason for this is unclear but could be due to a number of factors including differences between mouse lines, as has been observed for other modified capsid variants. While there was some variability, it was clear from this experiment that αFAP-AAV infected the xenografted tumor more efficiently than any other treatment.

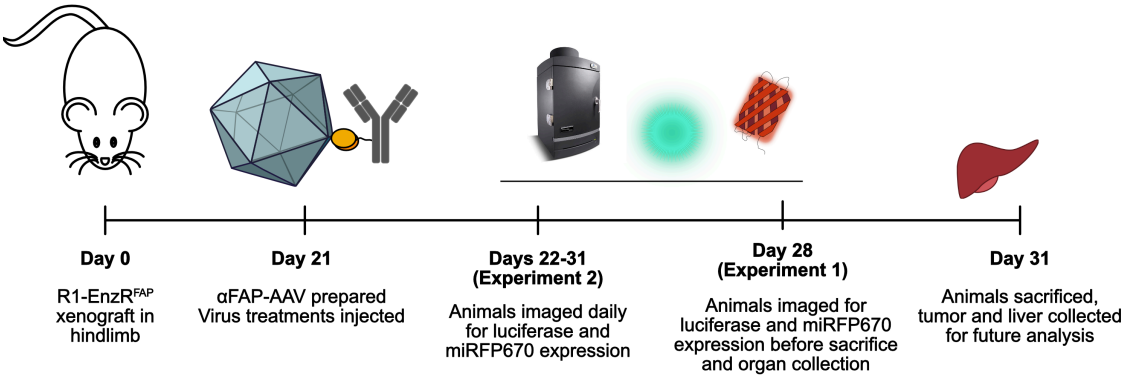
Results from the second experiment introduced more concern about the mechanism of infection. Like the first experiment, αFAP-AAV infected the tumor more effectively than any other treatment at the same MOI (MOI 1). Infection dropped significantly at lower MOI, which is not entirely surprising (**Figure 4.3b**). 10 days post-injection, he IgG-AAV infected the tumor more effectively than the base capsid not conjugated to any antibody. This suggests that there is some interaction between the tumor and either the IgG or αFAP antibody facilitating infection not via FAP antibody-antigen binding.

The low antibody-AAV conjugation efficiency was not unexpected and can be attributed to several factors. First, the reaction efficiency of mMobA is maximally

around 50% when the recombinantly produced protein is assayed in excess ssDNA[120]. Producing the enzyme as a chimera with a viral protein likely decreases this efficiency, as does conjugating the ssDNA to an antibody. The reaction of an antibody-ssDNA and AAV-mMobA is much more sterically hindered. The ratio of modified VP2 incorporated in the capsid is unknown, and based on other VP2 modification studies, is likely lower than the natural ratio (~three copies of VP2 per capsid) [17,44]. These limiting aspects could be improved in future work by changing the HUH tag used and titrating VP2 expression levels by changing its promoter. Importantly, the conjugation rate is still measurable and clearly still results in viral capsids that can infect cells via the antibody-antigen interaction.

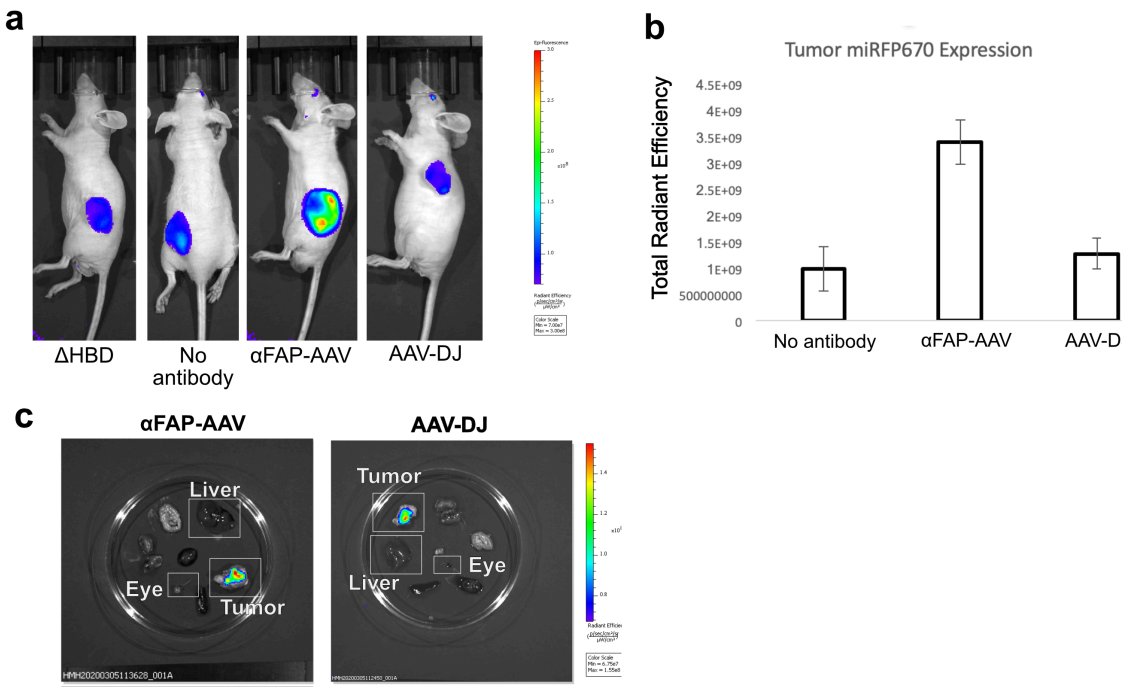
With these data in mind, the titers reported in this study and in past work are not representative of the number of antibody-conjugated, infective particles present. It is likely that a purified antibody-AAV would be several orders of magnitude more infective than unmodified serotypes. The mixture we are currently administering is effectively 96.2% filler with only 3.8% active virus that can infect cells via the antibody-antigen interaction. A more effective virus could lead to lower injection volumes, lower production cost, and fewer off-target effects. Future work will focus on improving purification to achieve a pure antibody-AAV solution.

Figure 4.1



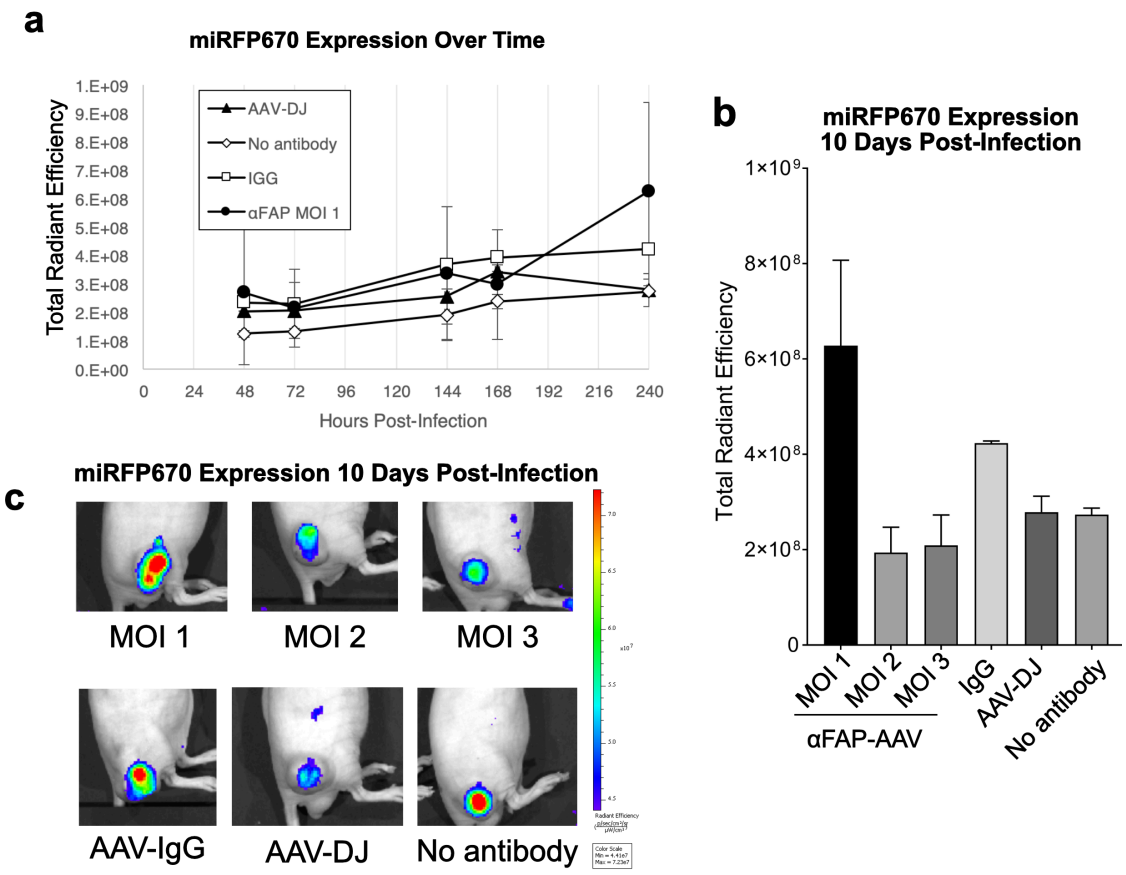
Experimental timeline for FAP+ prostate tumor targeting. Mice were subcutaneously injected with R1-EnzR^{FAP} cells to implant a hind leg tumor. Three weeks after xenograft, αFAP-AAV was prepared and injected into the tail vein of each mouse. In the first experiment, mice were imaged and organs harvested 7 days post-infection. Collected organs were imaged for miRFP670 expression. In the second experiment, mice were imaged at 48, 72, 144, 168, and 240 hours. At 240 hours (10 days), mice were scarified and organs were collected.

Figure 4.2



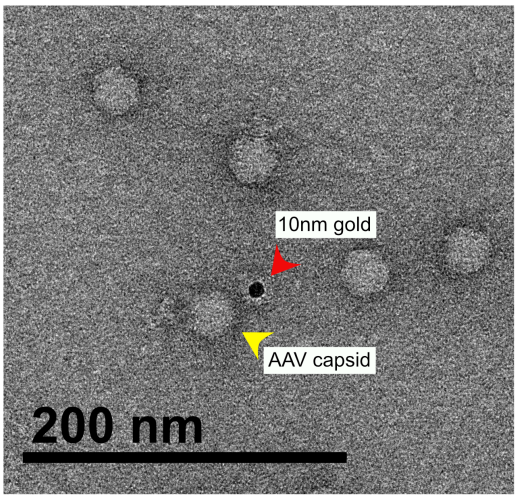
Analysis of mice seven days post-infection. A) Representative animal images from each experimental group. B) quantification of miRFP670 expression seven days post-injection.

Figure 4.3



Analysis of infected mice over a ten-day time course. A) Average miRFP670 expression measured as radiant efficiency 48, 72, 155, 168, and 240 hours after injection. B) Quantitation of miRFP670 expression 10 days post-infection C) Representative images of mice from each experimental group.

Figure 4.4



Conjugated	Full	9
	Empty	0
	Total	9
Unconjugated	Full	210
	Empty	15
	Total	225

Transmission electron microscopy images of antibody-AAV for calculation of conjugation efficiency. 6.4% of counted capsids were empty, 3.8% of capsids were conjugated to an antibody.

Chapter 5

Insertional mutagenesis libraries of adeno-associated virus to interrogate capsid allostery

This chapter is intended as a preliminary start for a future manuscript and as a reference for future project researchers

Zdechlik, A.C., He, Y., Nedrud, D., Coyote-Maestas, W.C., Hoffmann, M.D., Schmidt, D.A.

A.C.Z., W.C-M., and D.A.S. designed experiments. D.N. designed oligo pool libraries. M.D.H. assisted in preparation of the AAV-2 library and will carry out future experiments. A.C.Z. carried out all other experiments with input from W.C-M. and D.N. A.C.Z. authored the manuscript.

Preface

Adeno-associated virus has gained traction as a gene delivery vector in recent. However, its protein capsid remains largely uncharacterized. Using a domain insertion protocol recently developed by our lab (SPINE), we have begun a project to characterize AAV capsids with protein domains inserted at each position across all three capsid proteins (VP1-3) in two serotypes (AAV-DJ and AAV-2). While the results of this project are still outstanding, the work that has been completed so far is compiled here. The results of this project will provide information about the flexibility of the entire capsid and new candidate locations for insertion of functional domains. We also discuss the creation of a stable cell line to enable AAV library packaging with reliable genotype-phenotype linkages.

Introduction

Recombinant adeno-associated virus (AAV) is a promising vector for clinical gene delivery because of its small size and relatively low immunogenicity [7]. The two major challenges of using AAV in the clinic are its limited tropism and possible neutralization by preexisting antibodies [105,188]. Efforts to rationally engineer the capsid of AAV to have better tissue specificity or to evade the immune system have been ongoing since the 1990s [40]. While sometimes successful, directed mutations can prove a disastrous and tedious guess-and-check process [189]. Saturating mutagenesis is more likely to identify a productive mutant, as all possible positions have been tested [190]. In this work, we create a saturating domain insertion library from two serotypes of AAV to characterize capsid plasticity and locate regions permissible to functional domain insertion.

AAV is a non-enveloped virus composed of a 25nm diameter protein capsid surrounding a single-stranded (ss)DNA genome [24]. The protein capsid is made of three viral proteins named VP1, VP2, and VP3. The three proteins have overlapping coding sequences, with VP1 composed of the full 735 amino acid sequence, VP2 composed of amino acids 138-735, and VP3 composed of 203-735 [14,15]. VP3 is the most abundant component of the 60-subunit capsid; the ratio of VP1:VP2:VP3 is reported to be 1:1:10 [191,192]. At the time of assembly, both termini of all three proteins are contained within the capsid itself, making it challenging to create terminally attached protein chimeras [25]. Many viral protein engineering efforts for AAV focus on changing the tropism of the virus, typically

requiring the modified region to be readily displayed on the surface of the virus. This, combined with the fragility of the 60-mer, non-covalently assembled capsid has made directed engineering of AAV challenging.

In place of structure-guided engineering, high-throughput approaches, such as directed evolution and *in silico* analyses to characterize and manipulate the capsid of AAV have been very well received in the last ten years [52–54,193]. With increasingly accurate molecular imaging techniques such as cryo-EM, structural data is reaching a point of diminishing returns.

Our lab has recently developed a method of inserting a protein domain at each amino acid across a protein of interest [194]. This method is based on DNA synthesis, meaning that it is unbiased, unlike past library generation methods that relied upon transposase or DNase digests [195]. We wanted to use this method to interrogate the capsid of AAV, both to locate candidate regions for insertion of functional domains and to characterize permissibility of the capsid protein.

We chose to start by preparing insertion libraries for two serotypes, AAV-2 and AAV-DJ. AAV-DJ was selected from a shuffled capsid library of AAV-2, -4, -5, -8, and -9 to have high affinity for human hepatocytes and to be resistant to preexisting serum antibodies [39]. AAV-2 and AAV-DJ differ in only 60 of 735 amino acids but have notably different bioactivity, making them interesting to compare. For the AAV-DJ library, we chose to mutate the heparin binding domain to yield a virus that is effectively non-infective in cell culture as previously shown [78]. In the AAV-2 library, the heparin binding domain was left intact. All three viral protein start

codons were maintained in the AAV-2 library, meaning that an insertion would have to be tolerated in every protein of the viral capsid for a capsid to form and a genome to be packaged. In contrast, VP2 and VP3 start codons were mutated in the AAV-DJ library so only AAV-DJ VP1 would contain insertions. VP2 and VP3 were supplied in trans for library production. VP1 is unique among capsid proteins because it contains a phospholipase domain necessary for endosomal escape and genome expression [36,102,196]. Therefore, any capsid that results in productive infection can be assume to contain at least one copy of VP1.

AAV functionality is a complex term that can be broken down into many steps. The capsid must be able to assemble, package a genome, be stable enough for collection, bind the surface of a cell for infection and be successfully endocytosed, escape the endosome, and deliver its genome to the nucleus to be expressed. We will test each of these steps separately through carefully designed assays. To test assembly, virus will be produced in HEK293 cells via transient transfection. After virus collection, genomes will be amplified and sequenced. Any variants that appear in this analysis can assemble into capsids stable enough for virus production and collection. We will test for cell binding and endocytosis using a previously described protocol[197]. To test for genome expression, we will use fluorescence-activated cell sorting to isolate cells expressing mKate2, which is co-expressed with the VP library in the viral genome. We will then isolate and sequence genomes to determine which variants were collected.

One largely unaddressed problem in the field of AAV is the potential for cross-packaging of variants. In AAV libraries, the genotype-phenotype linkage is maintained by a given capsid protein variant packaging its corresponding coding sequence in the ssDNA genome [135]. Viral capsid assembly is stochastic and precedes genome packaging. If two variants are introduced to the same cell during production it is possible and likely that one protein variant would carry the ssDNA from the other. The way this has been addressed to this point is by lowering the amount of library plasmid DNA used to transfect packaging cells [117]. While this reduces the number of cross-packaged variants, it is not perfect and significantly reduces the titer of virus produced.

To address a similar problem of multiple library variants being expressed by a single transduced cell, the Fowler lab at the University of Washington has created a series of cell lines where expression is dependent upon integration [154]. Only a single integration even can occur in any cell, meaning that each cell can be analyzed as a single variant of the library. We have designed a version of this cell line suitable for AAV library production (**Figure 5.3**). In this system, a library staging plasmid is used for preparation of the AAV library with the serotype and inserted domain of interest. Notably, this construct only contains the terminal inverted terminal repeat (ITR) and lacks a promoter. When transfected into a cell without an available corresponding genomic landing pad, neither an AAV genome nor any capsid protein will be expressed from this plasmid. Expression of the AAV library and corresponding genome are dependent upon integration to the

customized AAV library genomic landing pad. This landing pad is based directly on the pLenti-TetBxb1-BFP-iCasp9-Blast (LLP-iCasp9-Blast) cell line designed by the Fowler lab [154]. Before integration of the AAV library from the staging plasmid, landing pad cells would express BFP and a synthetic form of Caspase 9 that causes cell death upon treatment with a small molecule (AP1903). The Caspase 9-BFP coding region is removed upon recombination. Cells would be treated with AP1903 after transfection with AAV library to remove any cells that had not experienced recombination.

In this system, as both AAV genome and capsid protein expression are dependent up on integration and there can be only one integrant per cell, we can assume perfect genotype-phenotype linkage. Helper plasmid and any required complementary viral protein plasmids could be transiently transfected into producer cells for viral library production. Because the iCasp9 system will have eliminated any nonintegrated cells, every cell can be assumed to be virus/variant-producing. As AAV library creation continues to gain popularity, it may be useful to create this cell line to enable researchers to maintain stringent genotype-phenotype linkages without the need for modified virus production conditions.

Methods & Materials

OLS design

Oligo pool sequences were designed as previously described using SPINE algorithm [195]. Briefly, the coding sequences of AAV-DJ and AAV-2 were cloned into a pATT plasmid backbone and mutated to remove all BsaI and BsmBI sites. For AAV-DJ, the heparin binding domain (HBD) was mutated to an HA tag and the start codons for VP2 and VP3 (T138, M203, M211, and M235) were mutated. For AAV-2, all VP start codons and the HBD were left intact. The 735AA (2205nt) VP1 CDS was divided into 14 fragments of 230 nucleotides each. At each nucleotide position, a pair of restriction sites was inserted *in silico*. For each fragment, unique subpool amplification barcodes were added to the ends. All AAV-DJ sequences were ordered pooled with several other libraries; AAV-2 sequences were similarly ordered at a later date.

Library preparation and validation

For each of the 14 fragments, backbone was amplified around the 230nt section used for the oligo pool. Oligo pool fragments were amplified using primers that bound to their subpool amplification identifying barcodes. All amplified products were purified on an agarose gel. The backbone and oligo for each fragment was assembled via 42-cycle Golden Gate reaction. Assembled reactions were digested with DpnI and PlasmidSafe (Lucigen) to remove contaminating template DNA. After digestion, each reaction was electroporated into competent

cells. Three dilutions were plated on antibiotic plates, and remaining transformed cells were expanded for 16 hours in 20 mL of media. Cells were pelleted and plasmid DNA was extracted using a miniprep kit (Zymo). At this point, an equimolar amount of DNA from each fragment was combined, and a sample was sent for NovaSeq sequencing to confirm equal coverage. Analytical digests and PCR were performed to interrogate coverage as well.

After pooling all fragments, the genetic handle was replaced with a chloramphenicol cassette using the same Golden Gate assembly workflow. Transformed cells were plated on ampicillin/chloramphenicol double antibiotic plates. This was done to remove contaminating non-library plasmid DNA. Plasmid DNA was prepared and the chloramphenicol cassette was replaced with the coding sequence for an anti-GFP nanobody via Golden Gate assembly. Finally, the nanobody-containing VP1 ORF was cloned into a plasmid backbone containing ITRs, a CMV promoter, and mKate2. At this point, another aliquot of the plasmid library was sent for NovaSeq S1 2x150bp sequencing.

Viral library generation

Prepared plasmid DNA for the VP1 library, a construct containing AAV2 rep promoter and VP2/3 ΔHBD-HA, and pHelper were transfected into 60% confluent HEK293-AAV cells. Cells were transfected using an established protocol for AAV library capsid production to minimize cross-packaging between variants [117]. 72 hours after transfection, cells were harvested and lysed using freeze/thaw cycles.

GFP affinity pulldown

GFP beads were prepared as previously described [78]. Briefly, GFP with a free C-terminal cysteine was recombinantly produced in *E. coli* and purified via Histidine affinity resin. Biotin-maleimide was conjugated to purified GFP protein to create biotin-GFP. Streptavidin magnetic beads (NEB) were incubated with excess biotin-GFP to create GFP beads. Excess GFP-biotin was removed by washes with low-salt buffer.

AAV-containing lysate was bound to GFP beads. Beads were pelleted using a magnet and supernatant was collected. GFP-beads were then washed three times with low-salt buffer to remove any weak non-GFP/nanobody interactions. After washing, beads were resuspended in PBS and treated with Benzonase (Sigma Aldrich), then proteinase to liberate viral genomes from bound capsids. Genomes were purified using a ssDNA clean-up kit (Zymo) and frozen a -20 for future sequencing.

Cloning for stable cell line

Plasmids pLenti-TetBxb1-BFP-2A-iCasp9-Blast(LLP-iCasp9-Blast) and AttB-EGFP(700) were a generous gift from Douglas Fowler's lab [154]. To prepare a staging construct for AAV libraries, a P2A site, mKate2, a poly adenylation signal, and a single ITR were cloned into pAttB_EGFP, replacing the eGFP. To prepare a lentiviral genome for creation of a cell line with an AAV landing pad, a single ITR

was cloned into pLenti-TetBxb1-BFP-2A-iCasp9-Blast(LLP-iCasp9-Blast) between the first LTR and Tet promoter. These plasmids have been sequenced and stored for future use.

Results, Discussion and Future Directions

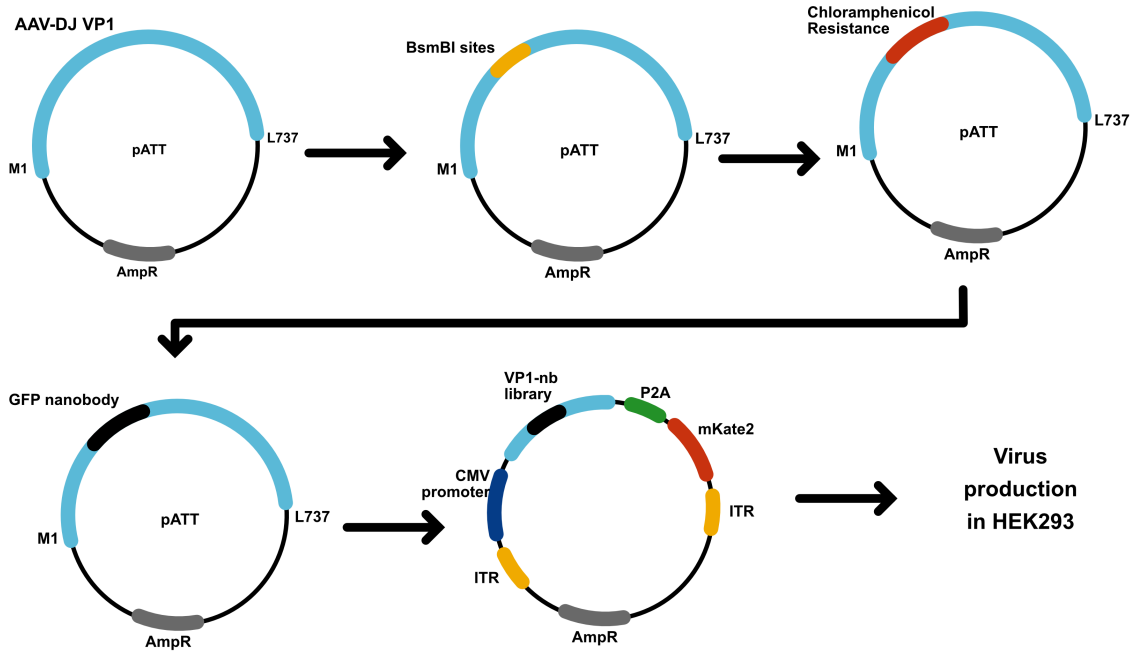
This project is still ongoing and few results can be reported at this time. OLS-mediated library generation of AAV-DJ VP1 and AAV-2 VP1/2/3 has been completed. From sequencing and digest analysis, the AAV-DJ VP1 library is minimally biased and has good coverage across all fragments (**Figure 5.2a, 5.2b**). This diversity was maintained through multiple cloning steps to the pAAV backbone used for virus production (**Figure 5.2c).** The AAV-2 library is currently being made and validated. The next step will be selecting new domains to test and performing functional assays on multiple domain insertion libraries. As aforementioned, we will test the library at each step throughout the production and infection process to characterize what impact insertion of each domain at each location has on the capsid. In the past, we have identified site/domain combinations that result in production of a capsid stable enough for collection and function of the domain, but interfere with infectivity. It is likely that modified capsid properties are determined by the unique combination of domain and insertion site, making it challenging to predict which combinations will produce a functional capsid without more experimental data.

Once this project is complete, we aim to have identified multiple sites on the capsid that are suitable for incorporation of new functional moieties, i.e. a site for an affinity tag that is compatible with VR-IV HUH tag insertion. Incorporation of new domains will increase the complexity of functions that can be performed by

the capsid; it could be possible to make an environment-sensitive capsid that is only infective within a specific tumor microenvironment[201, 202]. Increasing the complexity of AAV capsid function will result in more specific, safer delivery vectors[203].

Constructs for the AAV landing pad stable cell line have been prepared and are ready to be used to make lentivirus for cell line creation. Depending upon how much interest there is in AAV library approaches, this cell line could be worth producing in the future. One potential issue we could face is the efficiency of library production from a single genomic locus. In standard recombinant AAV production, viral proteins and genome are produced from multiple copies of plasmid DNA transfected into the cell. If expression is dependent upon genomic integration, all viral proteins and genomes will be expressed from that single locus. This is not dissimilar from how the natural virus is expressed from its integrated genome[204]. I hypothesize that proteins will also be expressed from free viral genome once it has been produced, making high-titer production likely. If viral production is too low in the stable cell line, we will reevaluate our strategy.

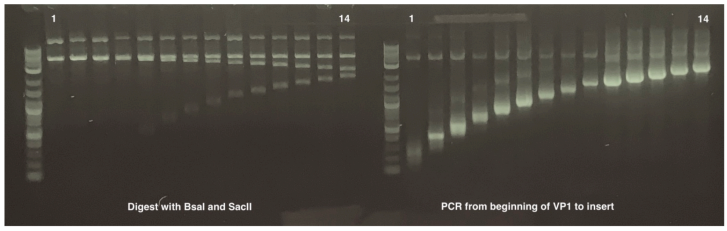
Figure 5.1



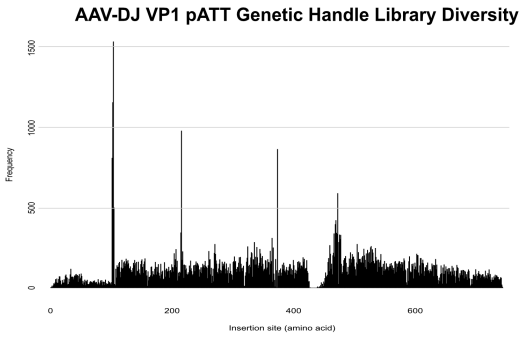
Workflow for creation of an AAV insertional scanning library. Initially, the library was assembled from the amplified oligo pool and VP1 in pATT staging backbone with a genetic handle as the inserted domain. The genetic handle was then exchanged for a chloramphenicol resistance cassette. The chloramphenicol cassette was then exchanged for the domain of interest, in this case, the anti-GFP nanobody. The VP1 CDS was then moved into an expression plasmid backbone containing a promoter and ITRs.

Figure 5.2

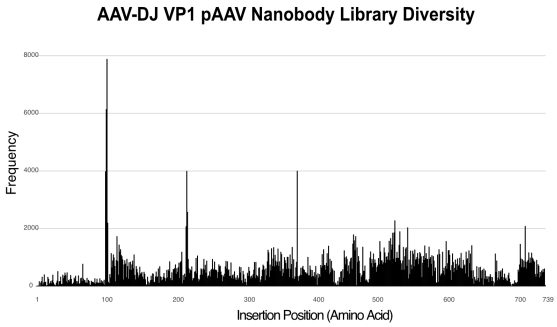
a



b

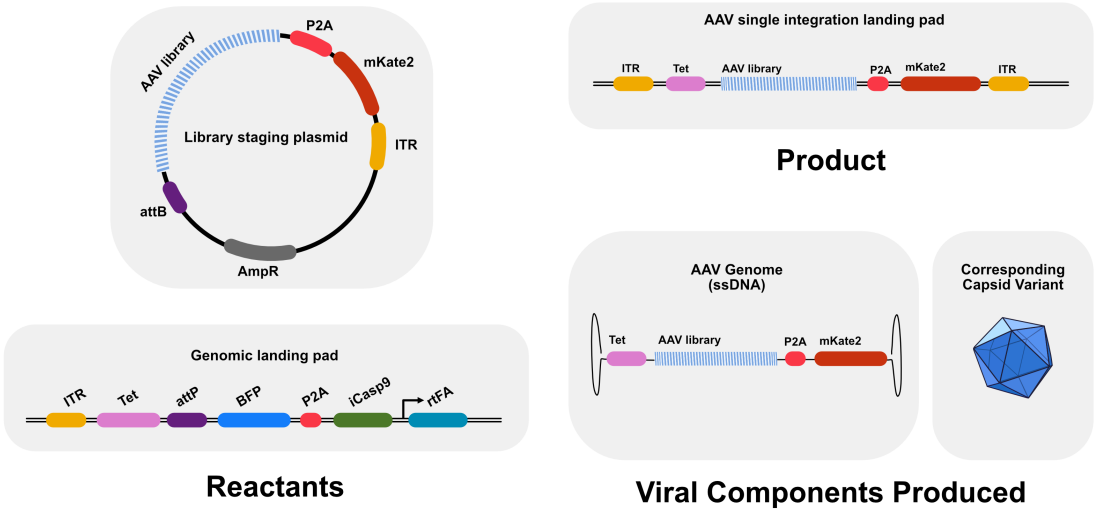


c



Analysis of AAV-DJ insertional library throughout generation. A) Analytical digest and PCR of fragments 1-14 of AAV-DJ VP1 insertional library in pATT backbone with genetic handle insert. B) Sequencing results from AAV-DJ VP1 library in pATT backbone with genetic handle insert, confirming results of a. c) Sequencing results from AAV-DJ VP1 library in pAAV backbone with GFP nanobody insert.

Figure 5.3



Design of a stable cell line for AAV library production without cross-packaging based on the single variant interrogation system described in [154].

Chapter 6

Conclusion & Future Directions

AAV-antibody conjugate engineering

I have created a modular system of attaching a targeting moiety to the capsid of AAV using a combination of protein engineering and chemical conjugation. These conjugates are effectively retargeted to infect cells via an antibody-antigen (or similar) interaction, losing natural tropism in favor of user-defined tropism. This addresses a long-time problem in the field and is a promising method of delivering genes specifically and to previously inaccessible targets. Future work will improve purification of antibody-conjugated capsids to produce a cleaner, better-characterized product. Conjugation efficiency could also be improved by testing different HUH tags, changing the ratio of modified protein to complementary capsid proteins, and changing reaction conditions.

I have also demonstrated that the variable region IV loop is a suitable location for nanobody and Gp2 insertion in VP1, VP2, and VP3. This is the first reported positon other than the heparin binding domain that can be modified across all three capsid proteins. All six of these modified capsids infected cells expressing their targeted antigen specifically and efficiently in culture. In the future, these constructs could serve as basis for inserting a targeting scaffold library into the capsid at the same position, and selecting for a tissue-specific variant *in vivo*. This project would be very similar to the work completed by the Gradinaru group where variants that effectively crossed the blood-brain-barrier were selected from a library

of capsids with a 7-amino acid diversified region [46,193]. Importantly, however, because existing AAV infectivity has been removed in our context, the selected targeting variant would be responsible for tissue binding. This means that the selected variant could be expressed independently from the capsid for biophysical characterization. This technique could be an update to current methods used for affinity selection, such as phage display or yeast display, and would account for the environment in which a targeting moiety is more likely to be used.

AAV-delivered prime editor

I assisted with creation and testing of an AAV-compatible version of prime editor. Editing rates are still low, and it will be important to demonstrate effective delivery using AAV in different contexts including primary cells and *in vivo*. It is worth noting that unmodified prime editor has been reported to have extremely variable efficiency, which seems to depend largely on editing context [172]. As prime editor guide design improves for *s. pyogenes* prime editor, it should provide information for how to improve editing efficiencies for *s. aureus* as well. The flexibility of prime editor makes it a candidate for treatment of many genetic disorders.

New editing tools are created at a breakneck pace, and it can be difficult to know which ones represent useful innovation. At this moment, safe, effective delivery is the main hurdle facing clinical gene editing, whether that delivery occurs *in vivo* or *ex vivo*[59]. Past editing tools have uniformly followed a path from initial

discovery, to modification for packaging in a delivery vector, to delivery *in vivo*[64,65,156]. Given this history, adapting prime editor for AAV-mediated delivery was the clear next step in bringing the technology one step closer to clinical trials.

Prostate tumor targeting

In this work, I demonstrated that the antibody-AAV technology described in my first paper can be safely and effectively used *in vivo*. While this data is still preliminary and will require further replication, it is a promising first step. The ability to deliver both a therapeutic effector (i.e. HSV-TK) and an imaging tool (miRFP670) fulfills the promise of creating effective, modern combination therapeutic/diagnostics. EM imaging of the capsid reinforces the belief that, while the vector is effective in its current state, further purification will be important to creating a maximally efficient gene delivery tool.

Cancer gene therapy using a targeted delivery vector is an important step towards safe and efficacious treatment. Many cancer-associated antigens have been described that are good targets for future development of antibody-directed gene delivery vectors [177,198]. In the future, we are interested in using the HUH-AAV technology to target other types of tumors, particularly in cancers where surgical removal is not an option. It would also be interesting to pursue *in vivo* generation of immunotherapy agents such as chimeric antigen receptors (CARs)[199]. In my mind, the clear next target for antibody-AAV mediated infection is bone marrow and blood cells, particularly CD34+ hematopoietic stem cells (HSCs). Delivery of

editing reagents specifically to HSCs represents a huge therapeutic potential for treatment of rare genetic blood and immune disorders (i.e. sickle cell disease, severe combined immunodeficiency) [60,200]. Whether this technology or another is used to achieve targeted delivery, gene therapy and/or editing for these disorders is incredibly promising.

Domain insertion library of AAV

This project is still in very early stages, and much of the data is yet to be collected. Thus far, I have prepared or assisted in preparation of domain insertion libraries for AAV-DJ and AAV-2. Libraries have been deep-sequenced and confirmed to be relatively unbiased. For this project, it will be important to analyze the insertion profile of many domains across the VP1 open reading frame. After collecting data for these two serotypes, we will investigate other AAV serotypes. In the future, this approach could be applied to other viral capsids and large multimeric proteins.

My design for an AAV library production stable cell line could be pursued by myself or others in the future to improve the fidelity of library genotype-phenotype linkage. The prevalence of cross-packaging in AAV libraries is a problem that has been minimally investigated, partially because it can be difficult to measure and partially because existing methods can select successful variants. It will be important to further characterize rates of cross-packaging for ideal AAV library production in the future.

Bibliography

1. Li, C. & Samulski, R. J. Engineering adeno-associated virus vectors for gene therapy. *Nat Rev Genet* **21**, 255-272 (2020).
2. van Haasteren, J., Li, J., Scheideler, O. J., Murthy, N. & Schaffer, D. V. The delivery challenge: fulfilling the promise of therapeutic genome editing. *Nat Biotechnol* **38**, 845-855 (2020).
3. Althoff, E. & Bulsara Speer, F. AveXis Announces Innovative Zolgensma Gene Therapy Access Programs for US Payers and Families. (2019).
4. Malone, D. C. et al. Cost-effectiveness analysis of using onasemnogene abeparvocec (AVXS-101) in spinal muscular atrophy type 1 patients. *J Mark Access Health Policy* **7**, 1601484 (2019).
5. Mingozzi, F. & High, K. A. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood* **122**, 23-36 (2013).
6. ATCHISON, R. W., CASTO, B. C. & HAMMON, W. M. ADENOVIRUS-ASSOCIATED DEFECTIVE VIRUS PARTICLES. *Science* **149**, 754-756 (1965).
7. Kotterman, M. A. & Schaffer, D. V. Engineering adeno-associated viruses for clinical gene therapy. *Nature Reviews Genetics* **15**, 445-451 (2014).
8. Sonntag, F., Schmidt, K. & Kleinschmidt, J. A. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proc Natl Acad Sci U S A* **107**, 10220-10225 (2010).
9. Smith, R. H. Adeno-associated virus integration: virus versus vector. *Gene Ther* **15**, 817-822 (2008).
10. Giraud, C., Winocour, E. & Berns, K. I. Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence. *Proc Natl Acad Sci U S A* **91**, 10039-10043 (1994).
11. Samulski, R. J. et al. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J* **10**, 3941-3950 (1991).
12. Hamilton, H., Gomos, J., Berns, K. I. & Falck-Pedersen, E. Adeno-associated virus site-specific integration and AAVS1 disruption. *J Virol* **78**, 7874-7882 (2004).
13. Papapetrou, E. P. & Schambach, A. Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Mol Ther* **24**, 678-684 (2016).
14. Muralidhar, S., Becerra, S. P. & Rose, J. A. Site-directed mutagenesis of adeno-associated virus type 2 structural protein initiation codons: effects on regulation of synthesis and biological activity. *J Virol* **68**, 170-176 (1994).
15. Becerra, S. P., Rose, J. A., Hardy, M., Baroudy, B. M. & Anderson, C. W. Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. *Proc Natl Acad Sci U S A* **82**, 7919-7923 (1985).
16. Girod, A. et al. The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. *J Gen Virol* **83**, 973-978 (2002).
17. Warrington, K. H. et al. Adeno-Associated Virus Type 2 VP2 Capsid Protein Is Nonessential and Can Tolerate Large Peptide Insertions at Its N Terminus. *Journal of Virology* **78**, 6595-6609 (2004).
18. Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. Production of high-titer helper-free retroviruses by transient transfection. *Proc Natl Acad Sci USA* **90**, 8392-8396 (1993).

19. Grimm, D., Kay, M. A. & Kleinschmidt, J. A. Helper virus-free, optically controllable, and two-plasmid-based production of adeno-associated virus vectors of serotypes 1 to 6. *Molecular Therapy* **7**, 839-850 (2003).
20. Xie, Q. et al. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc Natl Acad Sci U S A* **99**, 10405-10410 (2002).
21. Padron, E. et al. Structure of Adeno-Associated Virus Type 4. *Journal of Virology* **79**, 5047-5058 (2005).
22. Nam, H.-J. et al. Structure of Adeno-Associated Virus Serotype 8, a Gene Therapy Vector. *Journal of Virology* **81**, 12260-12271 (2007).
23. Ng, R. et al. Structural characterization of the dual glycan binding adeno-associated virus serotype 6. *J Virol* **84**, 12945-12957 (2010).
24. Lerch, T. et al. Structure of AAV-DJ, a Retargeted Gene Therapy Vector: Cryo-Electron Microscopy at 4.5 Å Resolution. *Structure* **20**, 1310-1320 (2012).
25. Bleker, S., Sonntag, F. & Kleinschmidt, J. A. Mutational analysis of narrow pores at the fivefold symmetry axes of adeno-associated virus type 2 capsids reveals a dual role in genome packaging and activation of phospholipase A2 activity. *J Virol* **79**, 2528-2540 (2005).
26. Rayaprolu, V. et al. Comparative Analysis of Adeno-Associated Virus Capsid Stability and Dynamics. *Journal of Virology* **87**, 13150-13160 (2013).
27. Myers, M. W. & Carter, B. J. Adeno-associated virus replication. The effect of L-canavanine or a helper virus mutation on accumulation of viral capsids and progeny single-stranded DNA. *J Biol Chem* **256**, 567-570 (1981).
28. Ling, C. et al. The Adeno-Associated Virus Genome Packaging Puzzle. *J Mol Genet Med* **9**, (2015).
29. Srivastava, A., Lusby, E. W. & Berns, K. I. Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J Virol* **45**, 555-564 (1983).
30. Chamberlain, K., Riyad, J. M. & Weber, T. Expressing Transgenes That Exceed the Packaging Capacity of Adeno-Associated Virus Capsids. *Human Gene Therapy Methods* **27**, 1-12 (2016).
31. Dong, J.-Y., Fan, P.-D. & Frizzell, R. A. Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. *Human Gene Therapy* **7**, 2101-2112 (1996).
32. Wu, Z., Yang, H. & Colosi, P. Effect of genome size on AAV vector packaging. *Mol Ther* **18**, 80-86 (2010).
33. Allocca, M. et al. Serotype-dependent packaging of large genes in adeno-associated viral vectors results in effective gene delivery in mice. *J Clin Invest* **118**, 1955-1964 (2008).
34. Kern, A. et al. Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. *Journal of Virology* **77**, 11072-11081 (2003).
35. Pillay, S. et al. An essential receptor for adeno-associated virus infection. *Nature* **530**, 108-112 (2016).
36. Stahnke, S. et al. Intrinsic phospholipase A2 activity of adeno-associated virus is involved in endosomal escape of incoming particles. *Virology* **409**, 77-83 (2011).
37. Kerr, J. R., Cotmore, S. F., Bloom, M. E., Linden, R. M. & Parrish, C. R. *Parvoviruses* (Hodder Arnold, London, 2006).
38. Ferrari, F. K., Samulski, T., Shenk, T. & Samulski, R. J. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J Virol* **70**, 3227-3234 (1996).

39. Grimm, D. et al. In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J Virol* **82**, 5887-5911 (2008).
40. Yang, Q. et al. Development of novel cell surface CD34-targeted recombinant adenoassociated virus vectors for gene therapy. *Hum Gene Ther* **9**, 1929-1937 (1998).
41. Hagen, S. et al. Modular adeno-associated virus (rAAV) vectors used for cellular virus-directed enzyme prodrug therapy. *Sci Rep* **4**, 3759 (2014).
42. Bouard, D., Alazard-Dany, D. & Cosset, F. L. Viral vectors: from virology to transgene expression. *Br J Pharmacol* **157**, 153-165 (2009).
43. Stumpp, M. T., Binz, H. K. & Amstutz, P. DARPins: a new generation of protein therapeutics. *Drug Discov Today* **13**, 695-701 (2008).
44. Münch, R. C. et al. Off-target-free gene delivery by affinity-purified receptor-targeted viral vectors. *Nature Communications* **6**, (2015).
45. Mao, Y. et al. Single point mutation in adeno-associated viral vectors -DJ capsid leads to improvement for gene delivery in vivo. *BMC Biotechnol* **16**, 1 (2016).
46. Deverman, B. E. et al. Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat Biotechnol* **34**, 204-209 (2016).
47. Hordeaux, J. et al. The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice. *Mol Ther* **26**, 664-668 (2018).
48. Hordeaux, J. et al. The GPI-Linked Protein LY6A Drives AAV-PHP.B Transport across the Blood-Brain Barrier. *Molecular Therapy* **27**, 912-921 (2019).
49. Huang, Q. et al. Delivering genes across the blood-brain barrier: LY6A, a novel cellular receptor for AAV-PHP.B capsids. *PLoS One* **14**, e0225206 (2019).
50. Upadhyay, G. Emerging Role of Lymphocyte Antigen-6 Family of Genes in Cancer and Immune Cells. *Front Immunol* **10**, 819 (2019).
51. Judd, J. et al. Random Insertion of mCherry Into VP3 Domain of Adeno-associated Virus Yields Fluorescent Capsids With no Loss of Infectivity. *Mol Ther Nucleic Acids* **1**, e54 (2012).
52. Ogden, P. J., Kelsic, E. D., Sinai, S. & Church, G. M. Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. *Science* **366**, 1139-1143 (2019).
53. Bryant, D. H. et al. Deep diversification of an AAV capsid protein by machine learning. *Nat Biotechnol* (2021).
54. Marques, A. D. et al. Applying machine learning to predict viral assembly for adeno-associated virus capsid libraries. *Mol Ther Methods Clin Dev* **20**, 276-286 (2021).
55. Fuller-Carter, P. I., Basiri, H., Harvey, A. R. & Carvalho, L. S. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun Biol* **1**, 54 (2018).
56. Gaudet, D. et al. Long-Term Retrospective Analysis of Gene Therapy with Alipogene Tiparvovec and Its Effect on Lipoprotein Lipase Deficiency-Induced Pancreatitis. *Human Gene Therapy* **27**, 916-925 (2016).
57. Senior, M. After Glybera's withdrawl, what's next for gene therapy? *Nature Biotechnology* **35**, 491-492 (2017).
58. Al-Zaidy, S. A. & Mendell, J. R. From Clinical Trials to Clinical Practice: Practical Considerations for Gene Replacement Therapy in SMA Type 1. *Pediatr Neurol* **100**, 3-11 (2019).
59. Rinde, M. Interview: Jennifer Doudna. *Distillations* (2019).

60. Frangoul, H. et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N Engl J Med* **384**, 252-260 (2021).
61. Kaiser, J. Gene therapy trials for sickle cell disease halted after two patients develop cancer. *Science* (2021).
62. Wright, A. V. et al. Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci U S A* **112**, 2984-2989 (2015).
63. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **33**, 139-142 (2015).
64. Levy, J. M. et al. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat Biomed Eng* **4**, 97-110 (2020).
65. Ran, F. A. et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191 (2015).
66. Villiger, L. et al. Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nat Med* **24**, 1519-1525 (2018).
67. Maeder, M. L. et al. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat Med* **25**, 229-233 (2019).
68. Tolmachov, O. E. Split vector systems for ultra-targeted gene delivery: a contrivance to achieve ethical assurance of somatic gene therapy in vivo. *Med Hypotheses* **83**, 211-216 (2014).
69. Kelly, E. & Russell, S. J. History of oncolytic viruses: genesis to genetic engineering. *Mol Ther* **15**, 651-659 (2007).
70. Lundstrom, K. New frontiers in oncolytic viruses: optimizing and selecting for virus strains with improved efficacy. *Biologics* **12**, 43-60 (2018).
71. Kamimura, K., Yokoo, T., Abe, H. & Terai, S. Gene Therapy for Liver Cancers: Current Status from Basic to Clinics. *Cancers (Basel)* **11**, (2019).
72. Xu, X. et al. Adeno-associated virus (AAV)-based gene therapy for glioblastoma. *Cancer Cell Int* **21**, 76 (2021).
73. Freytag, S. O., Stricker, H., Movsas, B. & Kim, J. H. Prostate cancer gene therapy clinical trials. *Mol Ther* **15**, 1042-1052 (2007).
74. Santiago-Ortiz, J. L. & Schaffer, D. V. Adeno-associated virus (AAV) vectors in cancer gene therapy. *J Control Release* **240**, 287-301 (2016).
75. Grifman, M. et al. Incorporation of tumor-targeting peptides into recombinant adeno-associated virus capsids. *Mol Ther* **3**, 964-975 (2001).
76. Shi, W. & Bartlett, J. S. RGD inclusion in VP3 provides adeno-associated virus type 2 (AAV2)-based vectors with a heparan sulfate-independent cell entry mechanism. *Molecular Therapy* **7**, 515-525 (2003).
77. Ried, M. U., Girod, A., Leike, K., Buning, H. & Hallek, M. Adeno-Associated Virus Capsids Displaying Immunoglobulin-Binding Domains Permit Antibody-Mediated Vector Retargeting to Specific Cell Surface Receptors. *Journal of Virology* **76**, 4559-4566 (2002).
78. Zdechlik, A. C., He, Y., Aird, E. J., Gordon, W. R. & Schmidt, D. Programmable Assembly of Adeno-Associated Virus-Antibody Composites for Receptor-Mediated Gene Delivery. *Bioconjug Chem* (2019).
79. Srivastava, A. In vivo tissue-tropism of adeno-associated viral vectors. *Curr Opin Virol* **21**, 75-80 (2016).
80. Harbison, C. E. et al. Examining the cross-reactivity and neutralization mechanisms of a panel of mAbs against adeno-associated virus serotypes 1 and 5. *J Gen Virol* **93**, 347-355 (2012).

81. Stachler, M. D. & Bartlett, J. S. Mosaic vectors comprised of modified AAV1 capsid proteins for efficient vector purification and targeting to vascular endothelial cells. *Gene Ther* **13**, 926-931 (2006).
82. Genemedi. AAV serotypes and AAV tissue-specific tropism. (2021).
83. Qing, K. et al. Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. *Nat Med* **5**, 71-77 (1999).
84. Summerford, C., Bartlett, J. S. & Samulski, R. J. AlphaVbeta5 integrin: a co-receptor for adeno-associated virus type 2 infection. *Nat Med* **5**, 78-82 (1999).
85. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Mol Ther* **16**, 1073-1080 (2008).
86. Cheng, B. et al. Development of optimized AAV3 serotype vectors: mechanism of high-efficiency transduction of human liver cancer cells. *Gene Ther* **19**, 375-384 (2012).
87. Vercauteren, K. et al. Superior In vivo Transduction of Human Hepatocytes Using Engineered AAV3 Capsid. *Mol Ther* **24**, 1042-1049 (2016).
88. Mietzsch, M., Broecker, F., Reinhardt, A., Seeberger, P. H. & Heilbronn, R. Differential adeno-associated virus serotype-specific interaction patterns with synthetic heparins and other glycans. *J Virol* **88**, 2991-3003 (2014).
89. Klein, R. L. et al. Efficient neuronal gene transfer with AAV8 leads to neurotoxic levels of tau or green fluorescent proteins. *Mol Ther* **13**, 517-527 (2006).
90. Balakrishnan, B. & Jayandharan, G. R. Basic biology of adeno-associated virus (AAV) vectors used in gene therapy. *Curr. Gene Ther.* **14**, 86-100 (2014).
91. Vandamme, C., Adjali, O. & Mingozzi, F. Unraveling the Complex Story of Immune Responses to AAV Vectors Trial After Trial. *Hum Gene Ther* **28**, 1061-1074 (2017).
92. Dismuke, D. J., Tenenbaum, L. & Samulski, R. J. Biosafety of recombinant adeno-associated virus vectors. *Curr. Gene Ther.* **13**, 434-452 (2014).
93. Betley, J. N. & Sternson, S. M. Adeno-associated viral vectors for mapping, monitoring, and manipulating neural circuits. *Hum Gene Ther* **22**, 669-677 (2011).
94. Murlidharan, G., Samulski, R. J. & Asokan, A. Biology of adeno-associated viral vectors in the central nervous system. *Front Mol Neurosci* **7**, 76 (2014).
95. Rutledge, E. A., Halbert, C. L. & Russell, D. W. Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. *J Virol* **72**, 309-319 (1998).
96. Agbandje-McKenna, M. & Kleinschmidt, J. AAV capsid structure and cell interactions. *Methods Mol Biol* **807**, 47-92 (2011).
97. Pillay, S. et al. Adeno-associated Virus (AAV) Serotypes Have Distinctive Interactions with Domains of the Cellular AAV Receptor. *J Virol* **91**, (2017).
98. Boutin, S. et al. Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors. *Hum. Gene Ther.* **21**, 704-712 (2010).
99. Wang, D. et al. Adeno-Associated Virus Neutralizing Antibodies in Large Animals and Their Impact on Brain Intraparenchymal Gene Transfer. *Mol Ther Methods Clin Dev* **11**, 65-72 (2018).
100. Asokan, A., Schaffer, D. V. & Jude Samulski, R. The AAV Vector Toolkit: Poised at the Clinical Crossroads. *Molecular Therapy* **20**, 699-708 (2012).
101. Grieger, J. C., Soltys, S. M. & Samulski, R. J. Production of Recombinant Adeno-associated Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest

of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector. *Molecular Therapy* **24**, 287-297 (2016).
102. Zádori, Z. et al. A viral phospholipase A2 is required for parvovirus infectivity. *Developmental Cell* **1**, 291-302 (2001).
103. Zinn, E. et al. In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. *Cell Rep* **12**, 1056-1068 (2015).
104. Girod, A. et al. Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. *Nat Med* **5**, 1052-1056 (1999).
105. Mietzsch, M., Pénzes, J. J. & Agbandje-McKenna, M. Twenty-Five Years of Structural Parvovirology. *Viruses* **11**, (2019).
106. Müller, O. J. et al. Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat Biotechnol* **21**, 1040-1046 (2003).
107. Michelfelder, S. et al. Vectors selected from adeno-associated viral display peptide libraries for leukemia cell-targeted cytotoxic gene therapy. *Exp. Hematol.* **35**, 1766-1776 (2007).
108. Körbelin, J. et al. Pulmonary Targeting of Adeno-associated Viral Vectors by Next-generation Sequencing-guided Screening of Random Capsid Displayed Peptide Libraries. *Mol Ther* **24**, 1050-1061 (2016).
109. Morabito, G. et al. AAV-PHP.B-Mediated Global-Scale Expression in the Mouse Nervous System Enables GBA1 Gene Therapy for Wide Protection from Synucleinopathy. *Mol Ther* **25**, 2727-2742 (2017).
110. Allen, W. E. et al. Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. *Neuron* **94**, 891-907.e6 (2017).
111. Challis, R. C. et al. Systemic AAV vectors for widespread and targeted gene delivery in rodents. *Nat Protoc* **14**, 379-414 (2019).
112. Jackson, K. L., Dayton, R. D., Deverman, B. E. & Klein, R. L. Better Targeting, Better Efficiency for Wide-Scale Neuronal Transduction with the Synapsin Promoter and AAV-PHP.B. *Frontiers in Molecular Neuroscience* **9**, (2016).
113. Hinderer, C. et al. Severe Toxicity in Nonhuman Primates and Piglets Following High-Dose Intravenous Administration of an Adeno-Associated Virus Vector Expressing Human SMN. *Hum Gene Ther* **29**, 285-298 (2018).
114. Matsuzaki, Y. et al. Intravenous administration of the adeno-associated virus-PHP.B capsid fails to upregulate transduction efficiency in the marmoset brain. *Neuroscience Letters* **665**, 182-188 (2018).
115. Lux, K. et al. Green Fluorescent Protein-Tagged Adeno-Associated Virus Particles Allow the Study of Cytosolic and Nuclear Trafficking. *Journal of Virology* **79**, 11776-11787 (2005).
116. Münch, R. C. et al. Displaying High-affinity Ligands on Adeno-associated Viral Vectors Enables Tumor Cell-specific and Safe Gene Transfer. *Molecular Therapy* **21**, 109-118 (2013).
117. Schmit, P. F. et al. Cross-packaging and capsid mosaic formation in multiplexed AAV libraries. *Molecular Therapy Methods & Clinical Development* **17**, 107-121 (2020).
118. Starovasnik, M. A., Braisted, A. C. & Wells, J. A. Structural mimicry of a native protein by a minimized binding domain. *Proc Natl Acad Sci U S A* **94**, 10080-10085 (1997).
119. Muik, A. et al. Covalent coupling of high-affinity ligands to the surface of viral vector particles by protein trans-splicing mediates cell type-specific gene transfer. *Biomaterials* **144**, 84-94 (2017).

120. Lovendahl, K. N., Hayward, A. N. & Gordon, W. R. Sequence-Directed Covalent Protein-DNA Linkages in a Single Step Using HUH-Tags. *J Am Chem Soc* **139**, 7030-7035 (2017).
121. Boucas, J. et al. Engineering adeno-associated virus serotype 2-based targeting vectors using a new insertion site-position 453-and single point mutations. *J Gene Med* **11**, 1103-1113 (2009).
122. Rothbauer, U. et al. A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics* **7**, 282-289 (2008).
123. Harmsen, M. M. & De Haard, H. J. Properties, production, and applications of camelid single-domain antibody fragments. *Applied Microbiology and Biotechnology* **77**, 13-22 (2007).
124. Rhee, J. M. et al. In vivo imaging and differential localization of lipid-modified GFP-variant fusions in embryonic stem cells and mice. *genesis* **44**, 202-218 (2006).
125. Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193-199 (1991).
126. Scallan, C. D. et al. Human immunoglobulin inhibits liver transduction by AAV vectors at low AAV2 neutralizing titers in SCID mice. *Blood* **107**, 1810-1817 (2006).
127. Kruziki, M., Bhatnagar, S., Woldring, D., Duong, V. & Hackel, B. A 45-Amino-Acid Scaffold Mined from the PDB for High-Affinity Ligand Engineering. *Chemistry & Biology* **22**, 946-956 (2015).
128. Chan, J. Y., Hackel, B. J. & Yee, D. Targeting Insulin Receptor in Breast Cancer Using Small Engineered Protein Scaffolds. *Molecular Cancer Therapeutics* **16**, 1324-1334 (2017).
129. Mullard, A. Maturing antibody–drug conjugate pipeline hits 30. *Nature Reviews Drug Discovery* **12**, 329-332 (2013).
130. Reichert, J. M., Rosensweig, C. J., Faden, L. B. & Dewitz, M. C. Monoclonal antibody successes in the clinic. *Nat Biotechnol* **23**, 1073-1078 (2005).
131. Sano, T., Smith, C. L. & Cantor, C. R. Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science* **5079**, 120-122 (1992).
132. Ramnarain, D. B. et al. Differential gene expression analysis reveals generation of an autocrine loop by a mutant epidermal growth factor receptor in glioma cells. *Cancer Res* **66**, 867-874 (2006).
133. Inoue, T., Swain, A., Nakanishi, Y. & Sugiyama, D. Multicolor analysis of cell surface marker of human leukemia cell lines using flow cytometry. *Anticancer Res* **34**, 4539-4550 (2014).
134. Maness, P. F. & Schachner, M. Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. *Nat Neurosci* **10**, 19-26 (2007).
135. Nonnenmacher, M., van Bakel, H., Hajjar, R. J. & Weber, T. High capsid-genome correlation facilitates creation of AAV libraries for directed evolution. *Mol Ther* **23**, 675-682 (2015).
136. Dudek, A. M. et al. An Alternate Route for Adeno-associated Virus (AAV) Entry Independent of AAV Receptor. *J Virol* **92**, (2018).
137. Gurda, B. L. et al. Mapping a neutralizing epitope onto the capsid of adeno-associated virus serotype 8. *J Virol* **86**, 7739-7751 (2012).
138. Bonifacino, J. S. & Traub, L. M. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu Rev Biochem* **72**, 395-447 (2003).

139. Traub, L. M. & Bonifacino, J. S. Cargo recognition in clathrin-mediated endocytosis. *Cold Spring Harb Perspect Biol* **5**, a016790 (2013).
140. Sassoon, I. & Blanc, V. Antibody-drug conjugate (ADC) clinical pipeline: a review. *Methods Mol Biol* **1045**, 1-27 (2013).
141. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
142. Konermann, S. et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472-476 (2013).
143. Kakarougkas, A. & Jeggo, P. A. DNA DSB repair pathway choice: an orchestrated handover mechanism. *Br J Radiol* **87**, 20130685 (2014).
144. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).
145. Kurt, I. C. et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nature Biotechnology* **39**, 41-46 (2021).
146. Zhao, D. et al. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol* **39**, 35-40 (2021).
147. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157 (2019).
148. Ding, Q. et al. Permanent Alteration of PCSK9 With In Vivo CRISPR-Cas9 Genome Editing. *Circulation Research* **115**, 488-492 (2014).
149. Dai, X. et al. One-step generation of modular CAR-T cells with AAV-Cpf1. *Nat Methods* **16**, 247-254 (2019).
150. Wang, D., Zhang, F. & Gao, G. CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors. *Cell* **181**, 136-150 (2020).
151. Stevens, A. J. et al. A promiscuous split intein with expanded protein engineering applications. *Proceedings of the National Academy of Sciences* **114**, 8538-8543 (2017).
152. Nishimasu, H. et al. Crystal Structure of Staphylococcus aureus Cas9. *Cell* **162**, 1113-1126 (2015).
153. Choi, J. G. et al. Lentivirus pre-packed with Cas9 protein for safer gene editing. *Gene Ther* **23**, 627-633 (2016).
154. Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res* **48**, e1 (2020).
155. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**, 224-226 (2019).
156. Truong, D.-J. J. et al. Development of an intein-mediated split–Cas9 system for gene therapy. *Nucleic Acids Research* **43**, 6450-6458 (2015).
157. Zettler, J., Schütz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Letters* **583**, 909-914 (2009).
158. Martin, A. S. et al. A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes. *Sci Rep* **9**, 497 (2019).
159. Yourik, P., Fuchs, R. T., Mabuchi, M., Curcuru, J. L. & Robb, G. B. Staphylococcus aureus Cas9 is a multiple-turnover enzyme. *RNA* **25**, 35-44 (2019).
160. Tang, X. et al. Plant Prime Editors Enable Precise Gene Editing in Rice Cells. *Molecular Plant* **13**, 667-670 (2020).

161. Liu, Y. et al. Efficient generation of mouse models with the prime editing system. *Cell Discov* **6**, 27 (2020).
162. Glaser, A., McColl, B. & Vadolas, J. GFP to BFP Conversion: A Versatile Assay for the Quantification of CRISPR/Cas9-mediated Genome Editing. *Molecular Therapy - Nucleic Acids* **5**, e334 (2016).
163. Gisler, S. et al. Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. *Nature Communications* **10**, (2019).
164. Lesueur, L. L., Mir, L. M. & André, F. M. Overcoming the Specific Toxicity of Large Plasmids Electrotransfer in Primary Cells In Vitro. *Mol Ther Nucleic Acids* **5**, e291 (2016).
165. Schellenberger, V. et al. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat Biotechnol* **27**, 1186-1190 (2009).
166. Amet, N., Lee, H.-F. & Shen, W.-C. Insertion of the Designed Helical Linker Led to Increased Expression of Tf-Based Fusion Proteins. *Pharmaceutical Research* **26**, 523-528 (2009).
167. Lin, Q. et al. Prime genome editing in rice and wheat. *Nature Biotechnology* **38**, 582-585 (2020).
168. Dixon, A. S. et al. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chemical Biology* **11**, 400-408 (2016).
169. Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature Biotechnology* (2020).
170. Huang, T. P. et al. Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat Biotechnol* **37**, 626-631 (2019).
171. Wang, A. S. et al. The Histone Chaperone FACT Induces Cas9 Multi-turnover Behavior and Modifies Genome Manipulation in Human Cells. *Molecular Cell* **79**, 221-233.e5 (2020).
172. Kim, H. K. et al. Predicting the efficiency of prime editing guide RNAs in human cells. *Nat Biotechnol* (2020).
173. Kleinstiver, B. P. et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature Biotechnology* **33**, 1293-1298 (2015).
174. Altwaijry, N., Somani, S. & Dufès, C. Targeted nonviral gene therapy in prostate cancer. *Int J Nanomedicine* **13**, 5753-5767 (2018).
175. Hintz, H. M. et al. Imaging Fibroblast Activation Protein Alpha Improves Diagnosis of Metastatic Prostate Cancer with Positron Emission Tomography. *Clin Cancer Res* (2020).
176. Cancer, I. A. F. R. O. All cancers fact sheet 2020. (2021).
177. Glumac, P. M. et al. Exploitation of CD133 for the Targeted Imaging of Lethal Prostate Cancer. *Clin Cancer Res* (2019).
178. Hintz, H. M., Cowan, A. E., Shapovalova, M. & LeBeau, A. M. Development of a Cross-Reactive Monoclonal Antibody for Detecting the Tumor Stroma. *Bioconjug Chem* **30**, 1466-1476 (2019).
179. Brennen, W. N., Isaacs, J. T. & Denmeade, S. R. Rationale behind targeting fibroblast activation protein-expressing carcinoma-associated fibroblasts as a novel chemotherapeutic strategy. *Mol Cancer Ther* **11**, 257-266 (2012).

180. Brennen, W. N., Rosen, D. M., Wang, H., Isaacs, J. T. & Denmeade, S. R. Targeting carcinoma-associated fibroblasts within the tumor stroma with a fibroblast activation protein-activated prodrug. *J Natl Cancer Inst* **104**, 1320-1334 (2012).
181. Szot, C. et al. Tumor stroma-targeted antibody-drug conjugate triggers localized anticancer drug release. *J Clin Invest* **128**, 2927-2943 (2018).
182. Li, C. et al. Clinical pharmacology of vc-MMAE antibody-drug conjugates in cancer patients: learning from eight first-in-human Phase 1 studies. *MAbs* **12**, 1699768 (2020).
183. Loganzo, F., Sung, M. & Gerber, H. P. Mechanisms of Resistance to Antibody-Drug Conjugates. *Mol Cancer Ther* **15**, 2825-2834 (2016).
184. Shariat, S. F. & Slawin, K. M. Gene therapy for prostate cancer. *Rev Urol* **2**, 81-87 (2000).
185. Amreddy, N. et al. Recent Advances in Nanoparticle-Based Cancer Drug and Gene Delivery. *Adv Cancer Res* **137**, 115-170 (2018).
186. Shcherbakova, D. M. et al. Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nature Communications* **7**, (2016).
187. Kanazawa, T. et al. Suicide gene therapy using AAV-HSVtk/ganciclovir in combination with irradiation results in regression of human head and neck cancer xenografts in nude mice. *Gene Ther* **10**, 51-58 (2003).
188. Calcedo, R., Vandenberghe, L. H., Gao, G., Lin, J. & Wilson, J. M. Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses. *J Infect Dis* **199**, 381-390 (2009).
189. Brannigan, J. A. & Wilkinson, A. J. Protein engineering 20 years on. *Nat Rev Mol Cell Biol* **3**, 964-970 (2002).
190. Balabanova, L., Golotin, V., Podvolotskaya, A. & Rasskazov, V. Genetically modified proteins: functional improvement and chimeragenesis. *Bioengineered* **6**, 262-274 (2015).
191. Johnson, F. B., Ozer, H. L. & Hoggan, M. D. Structural proteins of adenovirus-associated virus type 3. *J Virol* **8**, 860-863 (1971).
192. Becerra, S. P., Koczot, F., Fabisch, P. & Rose, J. A. Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript. *J Virol* **62**, 2745-2754 (1988).
193. Chan, K. Y. et al. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nature Neuroscience* **20**, 1172-1179 (2017).
194. Coyote-Maestas, W., He, Y., Myers, C. L. & Schmidt, D. Domain insertion permissibility-guided engineering of allostery in ion channels. *Nature Communications* **10**, (2019).
195. Coyote-Maestas, W., Nedrud, D., Okorafor, S., He, Y. & Schmidt, D. Targeted insertional mutagenesis libraries for deep domain insertion profiling. *Nucleic Acids Res* (2019).
196. Dorsch, S. et al. The VP1 Unique Region of Parvovirus B19 and Its Constituent Phospholipase A2-Like Activity. *Journal of Virology* **76**, 2014-2018 (2002).
197. Berry, G. & Tse, L. Virus Binding and Internalization Assay for Adeno-associated Virus. *BIO-PROTOCOL* **7**, (2017).
198. Peters, C. & Brown, S. Antibody-drug conjugates as novel anti-cancer chemotherapeutics. *Biosci Rep* **35**, (2015).
199. Agarwal, S. et al. In Vivo Generation of CAR T Cells Selectively in Human $CD4^+$ Lymphocytes. *Mol Ther* **28**, 1783-1794 (2020).

200. Cavazzana, M., Six, E., Lagresle-Peyrou, C., André-Schmutz, I. & Hacein-Bey-Abina, S. Gene Therapy for X-Linked Severe Combined Immunodeficiency: Where Do We Stand. *Hum Gene Ther* **27**, 108-116 (2016).
201. Guenther, C. M. et al. Protease-Activatable Adeno-Associated Virus Vector for Gene Delivery to Damaged Heart Tissue. *Molecular Therapy* **27**, 611-622 (2019).
202. Judd, J. et al. Tunable protease-activatable virus nanonodes. *ACS Nano* **8**, 4740-4746 (2014).
203. Brun, M. J., Gomez, E. J. & Suh, J. Stimulus-responsive viral vectors for controlled delivery of therapeutics. *J Control Release* **267**, 80-89 (2017).
204. Smith, R. H. Adeno-associated virus integration: virus versus vector. *Gene Ther* **15**, 817-822 (2008).